

Methoden der multivariaten Analyse nominal und- ordinalskalierter Daten auf der Basis mehrdimensionaler Kontingenztabellen: eine Einführung anhand ausgewählter Ergebnisse aus dem DFG-Projekt 'Bildungsbeteiligung und soziale Mobilität in preußischen Städten des 19. Jahrhunderts'

Sensch, Jürgen

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Sensch, J. (1987). Methoden der multivariaten Analyse nominal und- ordinalskalierter Daten auf der Basis mehrdimensionaler Kontingenztabellen: eine Einführung anhand ausgewählter Ergebnisse aus dem DFG-Projekt 'Bildungsbeteiligung und soziale Mobilität in preußischen Städten des 19. Jahrhunderts'. *Historical Social Research*, 12(2/3), 7-171. <https://doi.org/10.12759/hsr.12.1987.2/3.7-171>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more Information see:
<https://creativecommons.org/licenses/by/4.0>

METHODEN DER MULTIVARIATEN ANALYSE NOMINAL- UND ORDINALSKALierter
DATEN AUF DER BASIS MEHRDIMENSIONALER KONTINGENZTABELLEN

Eine Einführung anhand ausgewählter Ergebnisse aus dem DFG-Projekt
"Bildungsbeteiligung und soziale Mobilität in preußischen Städten
des 19. Jahrhunderts"

Jürgen Sensch (*)

Abstract: Statistical methods for multidimensional contingency table analysis: An introduction to categorical regression models with applications to be analysis of social mobility and educational participation in Germany within the 19. century.

This contribution examines a variety of issues within the context of discrete data analysis in the social sciences.

The first section reviews the methodological basis of log-linear models for the analysis of multidimensional contingency tables: maximum likelihood estimation, testing procedures, model selection and parameterization of models, which concern the mutual relationships among variables without explicit response (dependent variable).

The second article extends this framework to statistical models with explicit dependent variables: Individualized logistic regression and logit modeling for contingency tables in relation to log-linear models. In the third chapter we discuss an appropriate measure of total variation for multinomial data, developed as R^2 quantity which parallel the analogous statistics in multiple regression for quantitative data. The fourth chapter discuss simultaneous equation models with discrete data (path analysis). We show that there are important differences between the categorical case and the continuous-variable case and discuss the connections and limitations of recursive probability models within the context of log-linear models.

The purpose of the fifth chapter is to overcome the difficulty of too much log-linear parameters in complex models through a special application of association measures to the log-linear model.

Our purpose in the sixth chapter is to describe the basic elements of the statistical methodology of generalized linear models (GLIM). We discuss applications for logit and log-linear models within this methodology.

(*) Address all communications to: Jürgen Sensch, Universität Bielefeld, Fakultät für Soziologie, D-4800 Bielefeld.

In the seventh chapter we give a brief review about the "iterative proportional adjustment" of a sample frequency table when the marginal totals are known. This procedure solve the problem of decomposing mobility and educational Partizipation into structurel and circulatory components within an intertemporal comparison.

The last chapter examines several Computer programs for discrete data analysis.

Vorbemerkung

Der vorliegende Beitrag ist das Ergebnis einer methodischen Auseinandersetzung mit neueren Verfahren der multivariaten Analyse qualitativer Daten im Rahmen des DFG-Projektes "Bildungsbeteiligung und soziale Mobilität in preußischen Städten des 19. Jahrhunderts" (vgl. Lundgreen et al. 1987). Die Analyse des schichtspezifischen Bildungsverhaltens und die Analyse der Bedeutung von Familie und Schule im Statuszuweisungsprozeß (also für Mobilität und Berufsvererbung) im regionalen und zeitlichen Vergleich erfordert eine Auswahl adäquater statistischer Analysemethoden, um die komplexe Zusammenhangsstruktur zwischen den (nicht-metrischen) Merkmalen STADT, ZEIT, HERKUNFTSSCHICHT, SCHULBILDUNG und SOHNBERUF aufzuhellen. Als besonders geeignet erwies sich das log-lineare Modell zur multivariaten Analyse qualitativer Daten, eine Analysemethode, die in neueren Arbeiten zur Bildungssoziologie und Mobilitätsforschung bereits angewandt wird (vgl. u.a. Allerbeck/Storck 1980; Featherman/Hauser 1978; Mayer 1978; Müller 1981).

Ziel der vorliegenden Studie ist es, die Anwendung des log-linearen Modells auf eine symmetrische Fragestellung und des Logit-Modells auf eine asymmetrische Fragestellung in den Grundzügen darzustellen. Zur Veranschaulichung wird dabei immer die gleiche Frage als Beispiel herangezogen: "Wie unterscheiden sich Duisburg und Minden hinsichtlich des schichtspezifischen Bildungsverhaltens?"

Ferner wird das Ziel verfolgt, die relevante methodische Literatur (vor allem aus der Soziologie und der Ökonomie/Ökonometrie) zu Spezialfragen aufzuarbeiten, sofern sie für dieses Projekt von Bedeutung war. Wesentliche Gesichtspunkte werden anhand ausgewählter Beispiele konkretisiert.

Mein besonderer Dank gilt Herrn Prof. Dr. G. Arminger (Gesamthochschule Wuppertal), der mich bei der Anwendung notwendiger EDV-Programme angeleitet und beraten hat. Die im letzten Abschnitt diskutierte Anpassung von mehrdimensionalen (Stichproben-) Häufigkeitstabellen an vorgegebene Randverteilungen der Grundgesamtheit wäre schließlich ohne die Zusammenarbeit mit Herrn Arminger nicht möglich gewesen, der auch das hierfür notwendige EDV-Programm erstellte.

Einführung in die Problemstellung

Die allgemeine Forderung nach verstärktem Einsatz multivariater Verfahren in der empirischen Wirtschafts- und Sozialforschung läßt sich aus den materiellen Fragestellungen dieser Disziplinen ableiten: Dem Komplexitätsgrad sozialwissenschaftlicher Fragestellungen konnte mit der Analyse von Zusammenhängen oder Abhängigkeiten zwischen jeweils zwei Beobachtungsmerkmalen nur ungenügend Rechnung getragen werden. Sogenannte bivariate Analysen, die zwar sehr anschaulich sind, können etwa dann zu Fehlinterpretationen führen, wenn eine beobachtete zweidimensionale Beziehung durch ein drittes Merkmal (oder auch weiterer Faktoren) hervorgerufen oder in Art und Ausmaß mitgeprägt ist. Der Versuch, sich einen Überblick über multivariate Verfahren zu verschaffen macht deutlich, daß es sich hierbei um eine Vielzahl unterschiedlicher methodischer Ansätze handelt, wobei die Unterschiede sich primär aus den formalen Zielsetzungen der Verfahren (symmetrische Analyse mit explizit abhängigem Beobachtungsmerkmal oder asymmetrischer Analyse mit explizit abhängigem Merkmal) und dem Skalenniveau der zugrundeliegenden Merkmale ergeben. Gemeinsam ist allen klassischen Ansätzen der multivariaten Datenanalyse ihr Anwendungskontext: An jedem Merkmalsträger (z.B. ein Individuum) werden mehrere Variablen beobachtet und man betrachtet all diese Größen (oder zumindest eine Teilmenge davon) für eine Stichprobe von Merkmalsträgern gleichzeitig in der statistischen Analyse und wertet sie gemeinsam aus. Unter dem Begriff multivariater Verfahren auf der Basis von Individualdaten lassen sich die unterschiedlichsten Analysetechniken subsumieren:

- Regressionsanalyse: eine metrische abhängige Variable wird durch eine oder mehrere unabhängige metrische Merkmale erklärt;
- Varianzanalyse: eine metrische abhängige Variable wird durch eine oder mehrere unabhängige nominale/ordinale Merkmale erklärt;
- Kovarianzanalyse: eine metrische abhängige Variable wird sowohl durch metrische als auch durch nominale/ordinale unabhängige Merkmale erklärt;
- Faktorenanalyse: dient der Reduktion des Merkmalsraumes auf nur wenige zugrundeliegende Faktoren;
- Clusteranalyse: dient der Reduktion des Merkmalsträgerraumes, in dem die Untersuchungseinheiten anhand ausgewählter Merkmale in homogene Gruppen zusammengefaßt werden,

um nur die gebräuchlichsten Verfahren zu nennen, die heute zum klassischen Anwendungsrepertoire in den Wirtschafts- und Sozialwissenschaften gehören. Varianz- und Kovarianzanalyse lassen sich als Spezialfälle einer Regressionsanalyse mit sogenannten Dumray-Variablen (d.h. 0/1-kodierten Stellvertretervariablen für nominale/ordinale unabhängige Merkmale) darstellen (vgl. z. B. Holm 1979, Rochel 1983). Die (Regres-

sions-) Parameter der erstgenannten drei Ansätze können mit Hilfe der Methode der kleinsten Quadrate (OLS: ordinary least squares) auf der Basis der individuellen Merkmalsträger geschätzt werden. Betrachten wir lediglich diskrete Merkmale (d.h. nominale und/oder ordinale), dann finden in der empirischen Forschungspraxis im Rahmen der bivariaten Kontingenztabellenanalyse, bei denen jeweils zwei Merkmale miteinander kreuztabelliert werden, die unterschiedlichen Assoziationsmaße Verwendung (vgl. die Gegenüberstellung in Böhning et al. 1982: 102ff.).

Auch für die bivariate Kontingenztabellenanalyse gilt die Einschränkung: Fehlinterpretationen ergeben sich dann, wenn weitere Merkmale die Beziehung zwischen zwei Merkmalen beeinflussen. In diesen Fällen und generell für Analysen komplexerer Beziehungsstrukturen zwischen diskreten Merkmalen ist es notwendig, sich wiederum einer multivariaten Perspektive zu bedienen. Diese wird oftmals dadurch zu erreichen versucht, daß man Kreuztabellen nach weiteren Merkmalen aufspaltet und so zu einem differenzierteren, aber meist unübersichtlichen und in den Tendenzen schwer entzifferbaren Zahlenmaterial gelangt. Eine solche Darstellung, die mehr der Dokumentation als der eigentlichen Datenanalyse dient, kann mit Hilfe von multivariaten Verfahren für diskrete Merkmale, die vor allem in den siebziger Jahren Eingang in die empirische Sozialforschung fanden, wesentlich systematisiert und verfeinert werden.

Gemeint sind im Rahmen einer Analyse mehrdimensionaler Kontingenztabellen vor allem zwei Verfahrensansätze:

1. ein in der Literatur vor allem durch die Arbeiten von Goodman (1978) mit dem Etikett 'Goodman-Ansatz* oder 'log-lineare Modelle' versehenes symmetrisches Analyseverfahren, dessen Ziel es ist "...to describe the structural relationships among variables corresponding to the dimensions of the Table" (Fienberg 1980, S. 3);
2. ein unter dem Namen GSK-Ansatz bekannt gewordener Regressionsansatz, der eng mit der Dummy-Variablen-Regression verwandt ist und zur Gruppe der asymmetrischen multivariaten Verfahren zu zählen ist. Die Bezeichnung GSK-Ansatz folgt nach den Autoren Grizzle/Starmer/Koch (1969) (s. die einführende Darstellung von Kühler 1979).

Der wesentliche statistische Unterschied zwischen diesen Ansätzen liegt in dem jeweils verwendeten Schätzverfahren:

- ML (Maximum-Likelihood) - Schätzung im Rahmen der log-linearen Modelle;
- gewichtete Kleinste-Quadrate-Schätzung (WLS: Weighted Least Squares) im Rahmen des GSK-Ansatzes.

Beiden Ansätzen gemeinsam ist ein Wechsel der Analyseebene: Während bei den klassischen multivariaten Verfahren die Datenanalyse auf der Ebene der Merkmalsträger (d.h. Untersuchungseinheiten einer Stichprobe) selbst durchgeführt wird, erfolgt die multivariate Analyse diskreter Daten auf einem gruppenbezogenen Aggregations-

niveau ("sub-class level"). Gegenstand einer mehrdimensionalen Kontingenztabellenanalyse sind die Zellhäufigkeiten bzw. Häufigkeiten von sogenannten Subpopulationen, die durch systematisches Kombinieren der Ausprägungen von den jeweils in einer Analyse einbezogenen Merkmalen resultieren. Die Anzahl der Subpopulationen hängt von der Zahl der Ausprägungen der (diskreten) Gruppierungsmerkmalen ab. Sie sind somit die Untersuchungseinheiten einer multivariaten Kontingenztabellenanalyse. Mit der Aggregation von Individualdaten zu Häufigkeiten einer mehrdimensionalen Häufigkeitstabelle sind keinerlei Informationsverluste verbunden.

Das Ziel multivariater Verfahren zur Analyse mehrdimensionaler Kontingenztabellen ist die vereinfachte, auf wichtige Grundgrößen verdichtete Wiedergabe eines nicht überschaubaren Datenmaterials. Die Ausgangsdaten sollen auf wenige Größen reduziert werden, die dann Grundlage für substanzwissenschaftliche Aussagen sind. So gesehen stellen die statistischen Informationen in mehrdimensionalen Tabellen und die aus diesen herausgearbeiteten Zusammenhangsstrukturen zunächst rein deskriptive Informationen dar. Es geht im Rahmen der Deskription allerdings nicht nur in erster Linie um die formale Reduktion der Datenvielfalt, sondern um eine sachbezogene Kennzeichnung des empirischen Befundes anhand modellspezifischer Parameter.

Ein weiterer Gesichtspunkt der statistischen Analyse mehrdimensionaler Kontingenztabellen liegt in der inferentiellen Zielsetzung. Statistische Inferenz zielt darauf ab, aus einem Stichprobenbefund generellere Aussagen in Form von Allsätzen abzuleiten. Den Hintergrund bilden in den empirischen Wissenschaften insbesondere Hypothesen, die aufgestellt und an der Erfahrung durch Beobachtung überprüft werden. Will man also die auf der Grundlage eines Stichprobenmaterials erarbeiteten, zunächst deskriptiv gebrauchten Ergebnisse zu Erklärungszwecken verwenden, dann setzt dies voraus, daß die ermittelten Zusammenhangsstrukturen auch für andere Massen (z.B. der Grundgesamtheit, aus der die Stichprobe entstammt) ihre Gültigkeit nicht verlieren.

Die Zielsetzung "Deskription" und "Inferenz" lassen sich selbstverständlich auch in dieser allgemeinen Charakterisierung für die klassischen Analyseverfahren feststellen; sie sind jedoch im Rahmen der multivariaten Kontingenztabellenanalyse modelltheoretisch jeweils unterschiedlich konzipiert.

Nach dieser (zunächst verkürzten) Einordnung und Charakterisierung der Kontingenztabellenanalyse wollen wir im folgenden noch kurz auf den Stellenwert dieser statistischen Verfahren unter substanzwissenschaftlichen Gesichtspunkten des Projekts "Bildungsbeteiligung und soziale Mobilität" eingehen. Ziel des Projekts war u.a., auf der Grundlage eines umfangreichen Datensatzes drei leitende Fragestellungen zu beantworten:

1. FAMILIE und SCHULE: In welchem Grad hängt der Bildungsstatus einer Person (nach Schultypenwahl und Abgangsklasse operationalisiert) von ihrer sozialen Herkunft ab?

2. BERUFLICHE MOBILITÄT: Wie offen bzw. geschlossen war die Sozialstruktur der untersuchten Gesellschaft im 19. Jahrhundert, gemessen an dem Grad der Immobilität zwischen den Generationen?

3 . FAMILIE, SCHULE, BERUF: In welchem Grad determinieren die soziale Herkunftsschicht und das erreichte Bildungsniveau den erzielten Berufsstatuts?

Wie sieht das Übergangsmuster für Söhne einer bestimmten Herkunftsschicht in eine bestimmte soziale Zielgruppe, moderiert durch zuvor erreichte Bildungsränge, aus?

Bildungschancen und soziale Mobilitätsprozesse vollziehen sich in wandelnden Strukturen, und zwar in zweierlei Form: im zeitlichen wie im regionalen Vergleich. Die Analyse des Bildungsverhaltens und der sozialen Mobilität erfolgte dementsprechend zum einen komparativ für die in ihrer Sozialstruktur sehr unterschiedlichen Städte Minden und Duisburg. Eine weitere Dimension des Vergleichs zielt auf die zeitliche Veränderungen der Bildungschancen und der Mobilität innerhalb der jeweiligen Stadt. Die Ausdifferenzierung der drei Leitfragestellungen nach den Dimensionen STADT und ZEIT führt zu einer komplexen (mehrdimensionalen) Zusammenhangs- und Abhängigkeitsstruktur der jeweils in einem Untersuchungsschritt einbezogenen Merkmale (s. die Variablenübersicht auf S. 12). Die Zusammenstellung ausgewählter Merkmale macht deutlich, daß lediglich diskrete Variablen (in den meisten Fällen nur auf einer Nominalskala gemessen) vorliegend). Der Einsatz klassischer multivariater Verfahren zur Abhängigkeitsanalyse, die zumindest für die jeweils als abhängig betrachtete Variable metrisches Skalenniveau voraussetzen, verbietet sich somit zur Beantwortung der o.g. Fragestellungen. Unter statistisch-methodologischen wie inhaltlichen Gesichtspunkten können konventionelle Verfahren wie der CHI-QUADRAT-Unabhängigkeitstest oder die Berechnung von Assoziationsmaßen im Rahmen einer bivariaten Tabellenanalyse (die zwar dem niedrigen Skalenniveau genügen würden) dem Analyseproblem nicht gerecht werden, da bzgl. der einzelnen Fragestellungen die Dimensionen STADT und ZEIT explizit als unabhängige (Kontroll-) Variablen jeweils in die Analyse miteinbezogen werden müssen. Aus diesem Grunde reicht eine sukzessive Analyse der einzelnen Merkmale nicht aus, notwendig ist eine simultane Untersuchung von Merkmalskombinationen aller für die jeweilige Fragestellung relevanten Dimensionen und damit eine multivariate Analyse. Der Ausweg hieß daher, alternative Verfahren für den Fall nominaler Merkmale zu diskutieren, wobei die Entscheidung in dem Projekt aufgrund des Allgemeinheitsgrades zugunsten der log-linearen Tabellenanalyse (für die symmetrischen Fragestellungen) und der Logit-Analyse (für die asymmetrischen Fragestellungen) ausfiel.

Log-lineare bzw. Logit-Modelle sind in besonderer Weise geeignet, die komplexe Zusammenhangs- bzw. Abhängigkeitsstruktur zwischen Stadt (Sozialökologie), Zeit, sozialer Herkunftsschicht, Schulbildung und Berufsstatus aufzuhellen. Dieser Verfahrenstyp hat bereits seit geraumer Zeit Eingang in die Bildungssoziologie und Mobilitätsforschung gefunden (2).

Neben der Frage nach einem dem Datenniveau angemessenen multivariaten Analyseverfahren waren in dem Projekt zwei weitere Probleme zu lösen:

- (1) Anpassung von (Stichproben-) Zellenhäufigkeiten mehrdimensionaler Kontingenztabellen an die Randverteilung der Merkmale in der Grundgesamtheit.
- (2) Im Rahmen von Zeitvergleichen: Zerlegung der Bildungschancen und der Mobilität in zwei unterschiedliche Komponenten:
 - Mobilität (bzw. Chancen), welche allein durch Strukturveränderungen der Berufs- bzw. Bildungsstruktur erzwungen ist;
 - jene Unterschiede, welche nicht durch Strukturveränderungen zu erklären sind, sondern die Durchlässigkeit der Schichtstruktur einer Gesellschaft andeuten.

Hinter dem ersten Punkt verbirgt sich das Problem, daß Stichproben meist keine repräsentativen Teile der Grundgesamtheit sind. Auf das Projekt bezogen: Die Schüler der einzelnen städtischen Schulen lassen sich zwar schichtspezifisch fassen, aber da wir mehr Gymnasiasten als Volksschüler erheben konnten, sind Aussagen über das schichtspezifische Bildungsverhalten (die sog. Abstromprozentuierung) nicht möglich. Die Werte einer Tabellenprozentuierung hängen in hohem Maße von den Randverteilungen der betrachteten Merkmale ab. Nun war uns aber die Struktur der eindimensionalen Häufigkeitsverteilungen in der Grundgesamtheit bekannt. Mit Hilfe eines speziellen Verfahrens (dem sog. "iterative proportional fitting", kurz IPF, das auch dem log-linearen Modellansatz zugrunde liegt) ist es jedoch möglich, die beobachteten Zellenhäufigkeiten an die Duisburger bzw. Mindener Randverteilungen der Grundgesamtheit anzupassen, wobei die Zusammenhangsstruktur der tabellierten Merkmale zeitraumsspezifisch erhalten bleibt.

Die in dieser Weise "angepaßten" Kontingenztabellen erweitern die Analysemöglichkeiten entscheidend:

- Bivariate Tabellen können nunmehr horizontal im Sinne einer Abstromprozentuierung gelesen werden. Damit sind u.a. Fragen nach dem schichtenspezifischen Bildungs- und Berufschancen beantwortbar.
- Multivariate Rechnungen gewinnen jetzt die Möglichkeit, die zeitlichen Veränderungen genauer zu analysieren: zu unterscheiden zwischen Ungleichheitseffekten (bei gleicher Randverteilung) und den Effekten von sich ändernden Randverteilungsstrukturen.

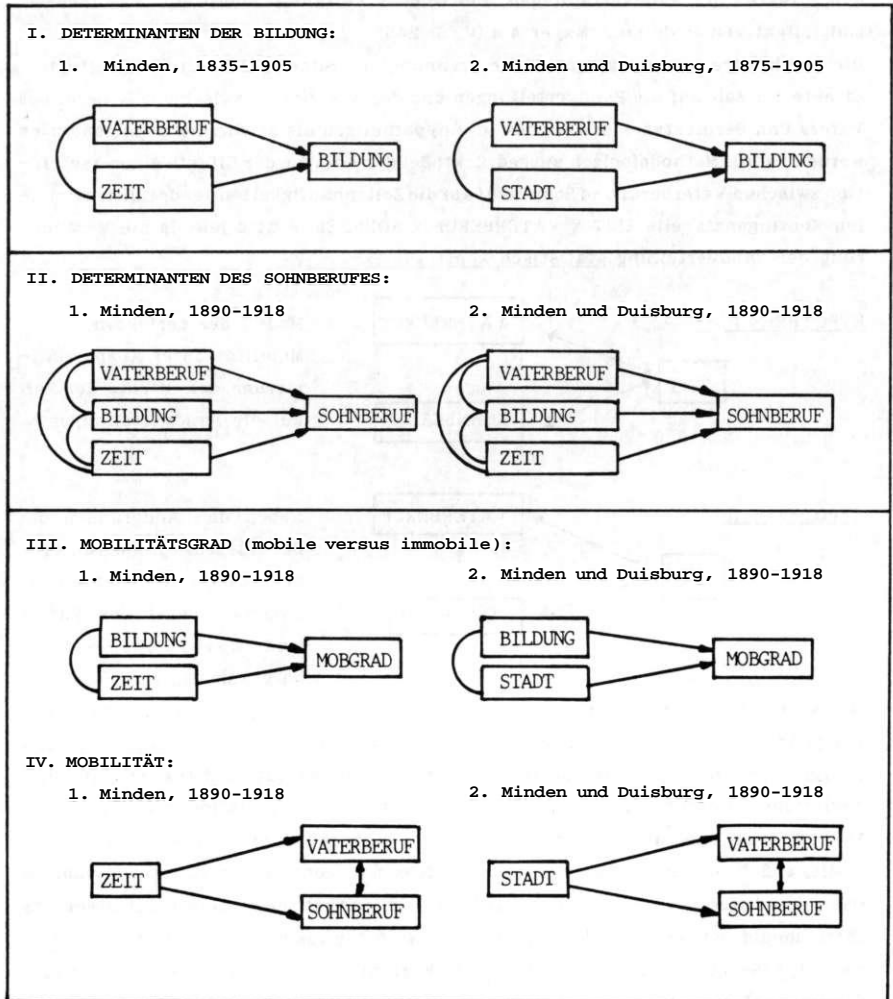
Klassifikation der Beobachtungsmerkmale für die mehrdimensionalen Kontingenzta- lenanalysen

Bildung als ab- hängige Variable	Sohnberuf als ab- hängige Variable
<u>Vaterberuf</u> 1 Besitzbürger 2 Bildungsbürger 3 alter Mittelstand 4 neuer Mittelstand 5 Arbeiter 6 untere Beamte und Angestellte	<u>Vaterberuf</u> 1 Oberschicht 2 alter Mittelstand 3 neuer Mittelstand 4 Unterschicht
<u>Bildung</u> 1 Volksschule 2 Mittelschule 3 Sexta bis Quarta 4 Tertia 5 Sekunda 6 Prima	<u>Bildung</u> 1 Volksschule 2 Mittelschule, Sexta bis Tertia 3 Sekunda, Prima
	<u>Sohnberuf</u> 1 Oberschicht 2 alter Mittelstand 3 neuer Mittelstand 4 Unterschicht
<u>Zeit</u> 1 Minden, 1835-1849 2 Minden, 1875-1885 3 Minden, 1895-1905	<u>Zeit</u> 1 Minden, 1890-1900 2 Minden, 1910-1918
<u>Stadt</u> 1 Minden, 1875-1905 2 Duisburg, 1875-1900	<u>Stadt</u> 1 Minden, 1890-1918 2 Duisburg, 1890-1913

Datenbasis

Kurz zusammengefaßt, läßt sich die Datenbasis, die im Prinzip mit Mikrozensusdaten vergleichbar ist, folgendermaßen charakterisieren: Für die beiden Städte, die seit 1850 rasant wachsende Gewerbe- und Handelsstadt Duisburg und die nur langsam wachsende Verwaltungsstadt Minden, sind je gut 4000 Schüler aller Schultypen aus archivalischen Quellen erfaßt und - über die Methode des "record linkage" - zu einer Kollektivbiographie zusammengestellt worden, die u.a. die oben aufgeführten Merkmale umfaßt.

Auf der Grundlage mehrdimensionaler Häufigkeitstabellen wurden z.B. folgende Kausalmodelle überprüft:

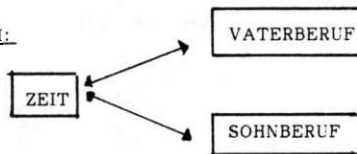


Der letzte Gesichtspunkt steht im unmittelbaren Zusammenhang mit Punkt (2). Angesprochen ist das Problem der Trennung strukturbedingter und strukturunabhängiger Austauschprozesse, das insbesondere in der Mobilitätsforschung breit diskutiert wurde

(vgl. u.a. Allerbeck/Storck (1980), S. 93ff; Mayer (1978), S. 237ff.). "Gesucht wurde also nach einem statistischen Modell, das es erlaubt, die Fluktuation im Sinne des Musters der Interaktion zwischen Herkunft und beruflicher Position zu isolieren von den Effekten der Randverteilungen. Ein solches Verfahren bieten die log-linearen, multiplikativen Modelle..." (Mayer a.a.O., S. 245).

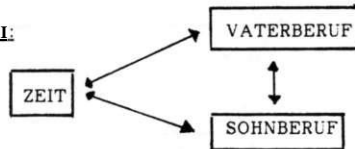
Die log-lineare Analyse gestattet die Trennung von reinen Randverteilungseffekten, Effekte der Zeit auf die Randverteilungen und der Assoziation zwischen z.B. Beruf des Vaters und Berufsstatus des Sohnes (die hypothetisch als zeitinvariant angenommen werden kann). Methodologisch ausgedrückt: Bei der Analyse der Effekte einer Assoziation zwischen Vaterberuf und Sohnberuf auf die Zellenhäufigkeiten der dreidimensionalen Kontingenztafel ZEIT x VATERBERUF x SOHNERUF wird jeweils die Veränderung der Randverteilung statistisch kontrolliert:

HYPOTHESE I:



Modell der perfekten Mobilität unter Auspartialisierung der Effekte der Zeit auf die Randverteilungen.

HYPOTHESE II:



Neben den Änderungen der Berufsstruktur wird eine zeitkonstante Beziehung zwischen Beruf des Vaters und Beruf des Sohnes berücksichtigt.

"It is now a widely accepted fact that by fitting a baseline model of statistical independence (perfect mobility to the father-son-table, (a) the structural mobility (marginal discrepancies) in the table is (are) partialled out, and (b) consequently, deviations from the independence can be equated with circulation mobility" (Sobel 1983, S. 721). Ohne auf den formalen Nachweis einzugehen, sei hier lediglich festgestellt, daß "...the model of independence does not 'control' for structural mobility (marginal differences) in any sense, and this implies that the association in a mobility table should not be equated (operationally) with the residual concept of circulation mobility" (Sobel a.a.O., S. 725). Mit anderen Worten: log-lineare Tabellenanalysen sind für eine Zerlegung der Mobilität in Struktur- und Zirkulationsmobilität nicht geeignet. Wir haben uns daher für einen alternativen Ansatz zur Trennung der Komponenten - Mobilität, die durch die jeweils gegebenen Randverteilungen erzwungen wird (Strukturmobilität) und

- Mobilität, die durch unterschiedliche Interaktionen innerhalb dieser Randverteilungen bedingt ist (Fluktuations- oder Zirkulationsmobilität) entschieden, der wiederum auf dem bereits genannten IPF-Verfahren beruht: Ausschaltung von Randverteilungseffekten durch erzwungene Anpassung beobachteter Kontingenztabellen an bestimmte, zeitlich verschobene Randverteilungen (zum Grundprinzip vgl. den Beitrag von Mosteller 1968). Subtrahiert man Abstrom- und Zustromprozente über die ursprüngliche und angepaßte Tabelle, werden Differenzen im Mobilitätsmuster sichtbar, die sich nicht mehr auf den Einfluß der unterschiedlichen Randverteilungen zurückführen lassen.

Aus den hier nur skizzenhaft umschriebenen Problemstellungen hat sich die vorliegende (im Kern methodische) Studie ergeben, die anhand eines - aus Gründen der Übersichtlichkeit - stark vereinfachten Projekt-Beispiels die einzelnen Lösungsschritte im Zusammenhang mit den jeweils zugrundeliegenden formal-statistischen Konzepten illustriert. Dem Beispiel liegt jeweils eine dreidimensionale Kontingenztafel zugrunde. Die ursprüngliche Zahl der Merkmalsausprägungen wurde jeweils durch Zusammenfassung von Kategorien reduziert, um ein hohes Maß an Anschaulichkeit (insbesondere bzgl. der Nachvollziehbarkeit von Rechenschritten) zu erzielen.

A.DIE LOG-LINEARE ANALYSE MEHRDIMENSIONALER KONTINGENZTABELLEN

Der formal-statistische Ansatz, den wir in unserer Untersuchung verwenden, ist das log-lineare Wahrscheinlichkeitsmodell zur Analyse mehrdimensionaler Häufigkeitstabellen. Da inzwischen mehrere Grundlagentexte und Lehrbücher zu dieser multivariaten Auswertungsmethode für nominale Variablen vorliegen (3), sei in der folgenden Darstellung nur soviel zu den methodischen Grundlagen der einzelnen Analyseschritten gesagt, wie zum Nachvollzug der im Hauptteil diskutierten Ergebnisse erforderlich scheint.

Die Akzente in diesem Abschnitt liegen auf der symmetrischen Fragestellung, bei der nicht zwischen abhängigen und unabhängigen Variablen unterschieden wird, während der Schwerpunkt des Logit-Modells im nächsten Abschnitt auf asymmetrischen Problemstellungen liegt. Zunächst wird dargestellt, wie sich die Analyse einer zweidimensionalen Häufigkeitstabelle im Rahmen der log-linearen Modelle gestaltet. Dabei werden die grundlegenden Konzepte entwickelt, die dann später für höherdimensionale Kontingenztabellen verallgemeinert werden.

Eine Kontingenztafel mit den Merkmalen A und B (z.B. Beruf des Vaters und Bildungsgrad des Sohnes) hat allgemein folgende Elemente:

Aufbau einer zweidimensionalen Kontingenztafel

B \ A	A1	A2	...	A _I	A _I	Σ
B1	n ₁₁	n ₂₁	...		n _{I1}	n _{.1}
B2	n ₁₂					
	⋮					
	⋮					
B _j				n _{Ij}		n _{.j}
B _J	n _{1J}	...			n _{IJ}	n _{.J}
Σ	n _{1.}			n _{1.}	n _{I.}	n _{..} = N

Diese Randsummen lassen sich mittels der Feldhäufigkeiten (n_{ij}) folgendermaßen ausdrücken:

$$n_{i.} = n_{ij} + n_{i2} + \dots + n_{iJ} = \sum_{j=1}^J n_{ij}$$

$$n_{.j} = n_{1j} + n_{2j} + \dots + n_{Ij} = \sum_{i=1}^I n_{ij}$$

Weiterhin ist die Gesamtzahl der Beobachtungen:

$$N = n_{..} = \sum_{i=1}^I \sum_{j=1}^J n_{ij} = \sum_{i=1}^I n_{i.} = \sum_{j=1}^J n_{.j}$$

Ein wesentliches Interesse bei der Analyse von beobachteten Feldhäufigkeiten besteht darin, stochastische Unabhängigkeiten bzw. Abhängigkeiten zwischen den jeweils betrachteten Merkmalen aufzudecken und diese zur Überprüfung von substanzwissenschaftlichen Hypothesen heranzuziehen.

Bezeichne nun für die beiden Merkmale A (z.B. Klassifikation nach dem Merkmal "Beruf des Vaters") und B (z.B. Klassifikation nach dem Merkmal "Bildungsrang des Sohnes")

$$p_{ij} = p_{ij} = P(A \text{ besitzt Ausprägung } A_i, B \text{ besitzt Ausprägung } B_j)$$

die gemeinsame Wahrscheinlichkeit des Auftretens der Merkmalskombination (A, B_j). Das Modell der hypothetisch angenommenen Unabhängigkeit läßt sich dann mit Hilfe des Multiplikationssatzes für Wahrscheinlichkeiten wie folgt aufstellen:

$$A.1 \quad p_{ij} = p_{i.} \cdot p_{.j} \quad , \quad i = 1, \dots, I \quad , \quad j = 1, \dots, J.$$

Die Nullhypothese (H_0 : die Zellenwahrscheinlichkeiten sind gleich dem Produkt der Randwahrscheinlichkeiten) legt somit Restriktionen für die gemeinsamen Wahrscheinlichkeiten des Auftretens der Merkmalskombination (A, B) fest. Unterstellt man diese Modellannahme, so ergeben sich die erwarteten Häufigkeiten in der Tabelle

$$A.2 \quad m_{ij} = E(N_{ij}) \quad m_{ij} = N (p_{i.} \cdot p_{.j})$$

Diese Modellgleichung ist durch theoretische (nicht beobachtete) Größen formuliert. In der Anwendung müssen für diese unbekannten Parameter erst Schätzwerte aus den beobachteten Stichprobendaten ermittelt werden. Abschließend ist die Gültigkeit des Modells auf der Grundlage der beobachteten Daten zu testen. Um Schätzer für ein bestimmtes Modell zu erhalten, ist die Stichprobenverteilung festzustellen, d.h. das stochastische Modell, das für die Entstehung der Daten unterstellt wird. Werden N Untersuchungseinheiten (vorgegebener Stichprobenumfang) nach zwei Merkmalen A und B in einer $(I \times J)$ -Kontingenztafel kreuzklassifiziert, so legt dieses Erhebungsschema eine Multinomialverteilung mit $I \times J$ Zellen und festem Stichprobenumfang N fest. Die gemeinsame Wahrscheinlichkeitsverteilung der beobachteten Häufigkeiten (n_{11}, \dots, n_{IJ}) ergibt eine Multinomialverteilung

$$A.3 \quad f(n_{11}, \dots, n_{IJ}) = \frac{N!}{n_{11}! \dots n_{IJ}!} \prod_{i,j} p_{ij}^{n_{ij}}$$

bzw. nach Umformung

$$A.4 \quad f(x) = \frac{N!}{n_{11}! \dots n_{IJ}!} \exp \left(\sum_{i,j} n_{ij} \ln p_{ij} - N \ln N \right)$$

Die Funktion A.4 nennt man auch die Likelihood-Funktion. Die Maximum-Likelihood-Schätzungen für das zweidimensionale Unabhängigkeitsmodell ergibt sich durch

$$A.5 \quad \hat{m}_{ij} = N(\hat{p}_{i.} \cdot \hat{p}_{.j}) = N \left(\frac{n_{i.}}{N} \cdot \frac{n_{.j}}{N} \right) = \frac{n_{i.}}{N} \cdot \frac{n_{.j}}{N}$$

d.h. m_{ij} läßt sich unmittelbar durch dies sogenannten "suffizienten" Randverteilungen $n_{i.}$ und $n_{.j}$ ausdrücken, die aus den n_{ij} Zellenhäufigkeiten ergeben. Damit läßt sich die Unabhängigkeitsthese (H_0 : $A \perp B$, d.h. A ist unabhängig von B) auch durch die Angabe der angepaßten (suffizienten) Randverteilungen ausdrücken:

$$H_0: A \perp B \quad (A) \quad (B)$$

Hat man für ein hypothetisches Modell die erwarteten Zellhäufigkeiten \hat{m}_{ij} berechnet, so ist zu prüfen, ob diese von den empirischen Häufigkeiten signifikant abweichen ($H_0: n_{ij} = \hat{m}_{ij}$). Diese Abweichungen berücksichtigt der von Pearson vorgeschlagene CHI-Quadrat-Test, der auf der CHI²-Statistik beruht:

$$A.6 \quad \chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(n_{ij} - \hat{m}_{ij})^2}{\hat{m}_{ij}}$$

Der CHI²-Wert ist bekanntlich, wenn die Nullhypothese nur zufälliger Abweichungen der erwarteten von den beobachteten Werten zutrifft, approximativ CHI²-verteilt mit den entsprechenden Freiheitsgraden. In der log-linearen Tabellenanalyse wird eine zu CHI² alternative Test-Statistik benutzt, die auch dem Vergleich der beobachteten Häufigkeiten mit den unter der Nullhypothese erwarteten Häufigkeiten dient:

$$A.7 \quad \chi_L^2 = \sum_{i=1}^I \sum_{j=1}^J 2n_{ij} \ln \frac{n_{ij}}{\hat{m}_{ij}}$$

Der Ausdruck χ_L^2 steht für "Likelihood Ratio CHI²". Wie χ^2 ist auch χ_L^2 asymptotisch CHI²-verteilt. Selbst wenn sich die Ergebnisse beider Tests im allgemeinen nur wenig unterscheiden, hat das Likelihood-Ratio-Kriterium den Vorteil der strukturellen Zerlegbarkeit; inhaltlich wird darauf noch im Rahmen der Sekundäranalyse eingegangen. Die Freiheitsgrade folgen analog zum χ^2 -Test.

Unter Berücksichtigung der Freiheitsgrade werden die $1-\alpha$ Prozentpunkte der χ^2 -Verteilung als Entscheidungsregel für den χ^2 -Test bei einem vorgegebenen Testniveau α benutzt. Diese Prozentpunkte gehen als kritische Werte (χ_{crit}^2) in den Test ein. Es wird bei $\chi^2 < \chi_{crit}^2$ bzw. $\chi_L^2 < \chi_{crit}^2$ die Hypothese (H_0) nicht abgelehnt, bei $\chi^2 \geq \chi_{crit}^2$ bzw. $\chi_L^2 \geq \chi_{crit}^2$ wird H_0 verworfen. Tabellen der Prozentpunkte der χ^2 -Verteilung für entsprechende Freiheitsgrade lassen sich in nahezu allen Statistikbüchern finden. In den weiteren Betrachtungen wird ein Test-Niveau von $\alpha = 0.05$ gewählt, demnach ist die Wahrscheinlichkeit 0.05, daß $\chi^2 \geq \chi_{crit}^2$ oder $\chi_L^2 \geq \chi_{crit}^2$ (Ablehnung von H_0), obwohl die Hypothese wahr ist. Diese Irrtumswahrscheinlichkeit, d.h. die Hypothese abzulehnen, obwohl sie zutrifft, bezeichnet man in der statistischen Testtheorie als Fehler 1. Art.

Wenn die Prüfgröße nicht signifikant ist ($p > 0.05$), können wir die Nullhypothese nicht zurückweisen; ein signifikanter Wert bedeutet demnach, daß das betrachtete Modell mit dem empirischen Befund nicht verträglich ist.

Bisher sind wir lediglich von einem Unabhängigkeitsmodell ausgegangen, das anhand der beobachteten Häufigkeiten mittels der χ^2 -Statistik auf Gültigkeit überprüft wird. Das log-lineare Modell sieht darüberhinaus die erwarteten Häufigkeiten unter H_0 i. F. von Regressionskoeffizienten an. Betrachten wir einmal den einfachsten Fall einer (2×2) -Tabelle (n_{11}, \dots, n_{22}) . Bei Gültigkeit der Unabhängigkeitshypothese

$$H_0: \hat{m}_{ij} = \frac{n_{i.} \cdot n_{.j}}{n_{..}}$$

erhält man durch Logarithmieren

$$A.8 \quad \ln \hat{m}_{ij} = \ln n_{i.} + \ln n_{.j} - \ln n_{..}$$

d.h. bei Gültigkeit der Unabhängigkeitshypothese läßt sich die zu erwartende Häufigkeit in einer Zelle darstellen als Summe dreier Terme, wobei der erste nur von der Zeile, der zweite nur von der Spalte und der dritte nur von dem Gesamtumfang der Stichprobe abhängt. Die allgemeine Struktur ist also von der Form

$$A.9 \quad \ln \hat{m}_{ij} = \mu + \mu_{A(i)} + \mu_{B(j)} \quad i, j = 1, 2$$

mit den Parametern μ , $\mu_{A(i)}$ und $\mu_{B(j)}$.

Betrachten wir nun die vier erwarteten Häufigkeiten als Meßwerte einer abhängigen Variablen für vier Untersuchungseinheiten und bilden wir als unabhängige Merkmale sogenannte Dummy-Variablen nach

$$A_1 \begin{cases} = 1 \text{ wenn das Merkmal A den Wert 1 hat,} \\ 0 \text{ sonst} \end{cases}$$

$$A_2 \begin{cases} = 1 \text{ wenn das Merkmal A den Wert 2 hat,} \\ 0 \text{ sonst} \end{cases}$$

$$B_1 \begin{cases} = 1 \text{ wenn das Merkmal B den Wert 1 hat,} \\ 0 \text{ sonst} \end{cases}$$

$$B_2 \begin{cases} = 1 \text{ wenn das Merkmal B den Wert 2 hat,} \\ 0 \text{ sonst} \end{cases}$$

2 x 2 Kontingenztabelle:

		B		
		1	2	
A	1	n_{11}	n_{12}	n_{1+}
	2	n_{21}	n_{22}	n_{2+}
		n_{+1}	n_{+2}	N

als Indikatoren für die Ausprägungen der Merkmale A und B, so können wir folgenden Regressionsansatz für die modellimplizierten Häufigkeiten des Unabhängigkeitsmodells entsprechend der Modellgleichung A.9 formulieren:

$$\begin{pmatrix} \ln \hat{m}_{11} \\ \ln \hat{m}_{12} \\ \ln \hat{m}_{21} \\ \ln \hat{m}_{22} \end{pmatrix} = \beta_0 \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \beta_1^A \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + \beta_2^A \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} + \beta_1^B \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} + \beta_2^B \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}.$$

In den Spalten stehen jeweils die Ausprägungen der Dummy-Variablen; über jede Spalte steht die zugehörige Kategorie der einbezogenen Merkmale, für die eine Stellvertreter-Variable gebildet wurde. Jede Zeile der Datenmatrix enthält somit einen Wert $\ln \hat{m}_{ij}$ als abhängige Größe und die Werte der Dummy-Variablen A_1 , A_2 , B_1 und B_2 , sowie einen konstanten Term als unabhängige Größen. Die fünf Spalten können nun in einer sog. Designmatrix \underline{X} zusammengefaßt werden:

$$\text{Designmatrix } \underline{X} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

(Wir verwenden den Begriff "Designmatrix" immer dann, wenn die Matrix \underline{X} ausschließlich aus Dummy-Variablen - einschließlich dem Absolutglied - besteht).

Ein Blick auf die Designmatrix zeigt, daß das Gleichungssystem des Regressionsansatzes so nicht lösbar ist, denn es bestehen lineare Abhängigkeiten zwischen den erklärenden Merkmalen (Spalten): Spalte 1 entspricht der Summe der Spalte 2, 3 und der Summe der Spalten 4, 5. (5) Damit die Matrix \underline{X} von vollem Spaltenrang ist (d.h. der Regressionsansatz lösbar wird), müssen wir lineare Restriktionen einführen, die auch als Reparametrisierungsbedingungen bezeichnet werden. Die in der log-linearen Analyse verwendeten Restriktionen der Regressionskoeffizienten β_1 , β_2 , β_3 und β_4 ist die Forderung

$$\begin{aligned} \text{A.10} \quad & \beta_1^A + \beta_2^A = 0 \quad \text{und} \quad \beta_3^B + \beta_4^B = 0 \\ \text{d.h.} \quad & \beta_2^A = -\beta_1^A \quad \text{und} \quad \beta_4^B = -\beta_3^B \end{aligned}$$

Berücksichtigt man diese üblicherweise in der Varianzanalyse verwendeten Nebenbedingungen

$$\boxed{\text{Summe der Effekte aller Dummy-Variablen eines Merkmals} = 0}$$

erhält man für das oben dargestellte Regressionsmodell die Gleichung

$$\begin{pmatrix} \ln \hat{m}_{11} \\ \ln \hat{m}_{12} \\ \ln \hat{m}_{21} \\ \ln \hat{m}_{22} \end{pmatrix} = \beta_0 \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \beta_1^A \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix} + \beta_3^B \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}, \quad \underline{X} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{bmatrix} \quad (\underline{X} \text{ hat nun vollen Spaltenrang})$$

$$\begin{bmatrix} \ln \hat{m}_{11} \\ \ln \hat{m}_{12} \\ \ln \hat{m}_{21} \\ \ln \hat{m}_{22} \end{bmatrix} = \beta_0 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \beta_i^A \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix} + \beta_j^B \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix} \quad \underline{X} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{bmatrix}$$

(\underline{X} hat nun vollen Spaltenrang).

Für das hier vorgestellte Modell ergibt sich wegen der Nebenbedingung gemäß (A.10) bei der Konstruktion der Designmatrix \underline{X} gerade die Effektkodierung, wenn die logarithmierten zu erwartenden Zellenhäufigkeiten als abhängig betrachtet werden. Der mit jeder Spalte der Designmatrix \underline{X} korrespondierende β -Effekt wird in der log-linearen Analyse mit "Lambda" (λ) bezeichnet. Man erhält so das log-lineare Unabhängigkeitsmodell, bestehend aus der linearen Parameterdarstellung zusammen mit den diskutierten Nebenbedingungen (A.10) durch

$$A.11 \quad \ln \hat{m}_{ij} = \lambda + \lambda_i^A + \lambda_j^B, \quad \text{wobei} \quad \sum_{i=1}^I \lambda_i^A = 0, \quad \sum_{j=1}^J \lambda_j^B = 0$$

gilt.

Die Lambda-Parameter des log-linearen Unabhängigkeitsmodells für den zweidimensionalen Fall ergeben sich unmittelbar aus (A.11) durch Summenbildung unter Berücksichtigung der Restriktionen zu:

Parameter des Unabhängigkeits-
Modells:

Geschätzte Parameter des log-linearen
Modells:

$$\begin{aligned} \lambda &= \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J \ln m_{ij} & \hat{\lambda} &= \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J \ln \left(\frac{n_{i.} n_{.j}}{N} \right) \\ \lambda_i^A &= \frac{1}{J} \left(\sum_{j=1}^J \ln m_{ij} \right) - \lambda & \hat{\lambda}_i^A &= \frac{1}{J} \left(\sum_{j=1}^J \ln \left(\frac{n_{i.} n_{.j}}{N} \right) \right) - \hat{\lambda} \\ \lambda_j^B &= \frac{1}{I} \left(\sum_{i=1}^I \ln m_{ij} \right) - \lambda & \hat{\lambda}_j^B &= \frac{1}{I} \left(\sum_{i=1}^I \ln \left(\frac{n_{i.} n_{.j}}{N} \right) \right) - \hat{\lambda} \end{aligned}$$

Wie diese Formeln zeigen, stellt (λ) die mittlere logarithmierte zu erwartende Zellenhäufigkeit dar, die Effekte werden dann als Abweichungen von diesem Mittelwert berechnet: Der Parameter λ_i^A stellt die mittlere Abweichung vom Gesamtmittelwert für die Ausprägung A_i dar und λ_j^B die mittlere Abweichung für die logarithmierte (erwartete) Häufigkeit der Ausprägung B_j .

Beschreibt ein Modell mit ausschließlich Ein-Faktoren-Effekten (= Haupteffekte) die beobachteten Daten nur unzureichend (d.h. $p < 0,05$), muß das log-lineare Modell

(A.11) verworfen werden. In diesem Fall gehen wir von einer Abhängigkeit der Merkmale A und B aus; es liegt dann nahe, durch Einführung zusätzlicher Parameter das Unabhängigkeitsmodell so zu erweitern, daß auch die Abhängigkeit zwischen den Merkmalen erfaßt werden kann. Das vollständige oder "saturierte" zweidimensionale log-lineare Modell ist gegeben durch

$$A.12 \quad \ln \hat{m}_{ij} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB} \quad \text{mit den Nebenbedingungen}$$

$$A.13 \quad \sum_i \lambda_i^A = \sum_i \lambda_j^B = \sum_i \lambda_{ij}^{AB} = \sum_j \lambda_{ij}^{AB} = 0$$

Der hinzukommende "Interaktionsparameter" λ_{ij}^{AB} ergibt sich aus der Modellgleichung (A.12) mit

$$\lambda_{ij}^{AB} = \ln \hat{m}_{ij} - \left(\lambda + \lambda_i^A + \lambda_j^B \right),$$

also als die Abweichung der $\ln \hat{m}_{ij}$ von dem durch die Parameter λ_i^A, λ_j^B erfaßten Anteile.

Wie man leicht zeigen kann, ist die Anzahl der Parameter des saturierten Modells genau gleich der Anzahl aller Zellen der Kontingenztafel. Es gelingt daher in diesem Fall eine vollständige Anpassung des korrespondierenden log-linearen Modells an die beobachteten Zellhäufigkeiten (d.h. $n_{ij} = \hat{m}_{ij}$), das in diesem Fall lediglich den empirischen Befund auf andere Weise ausdrückt.

Betrachten wir nochmals den einfachen Fall einer (2 x 2)-Tabelle: Die Parameterrestriktionen (A.13) implizieren, daß in Gleichung A.12

$$\begin{aligned} \lambda_2^A &= -\lambda_1^A, \\ \lambda_2^B &= -\lambda_1^B, \\ \lambda_{12}^{AB} &= \lambda_{21}^{AB} = -\lambda_{11}^{AB} \quad \text{und} \quad \lambda_{22}^{AB} = \lambda_{11}^{AB} \end{aligned} \quad \text{gilt.}$$

Die Gleichung A.12 ist schließlich äquivalent mit den Gleichungen

$$\begin{aligned} \ln(\hat{m}_{11}) &= \lambda + \lambda_1^A + \lambda_1^B + \lambda_{11}^{AB} \\ \ln(\hat{m}_{21}) &= \lambda - \lambda_1^A + \lambda_1^B - \lambda_{11}^{AB} \\ \ln(\hat{m}_{12}) &= \lambda + \lambda_1^A - \lambda_1^B - \lambda_{11}^{AB} \\ \ln(\hat{m}_{22}) &= \lambda - \lambda_1^A - \lambda_1^B + \lambda_{11}^{AB} \end{aligned}$$

wobei in dem dargestellten saturierten Modell $\hat{m}_{ij} = n_{ij}$ ist. Analog zur Varianzanalyse läßt sich Gleichung A.12 mit Hilfe von Kodiervariablen unter Annahme der Effekt-Kodierung als Spalten einer Designmatrix \underline{X} wie folgt übersetzen:

$\underline{Y} = \underline{XB}$, wobei

$$\underline{Y} = \begin{bmatrix} \ln(\hat{m}_{11}) \\ \ln(\hat{m}_{21}) \\ \ln(\hat{m}_{12}) \\ \ln(\hat{m}_{22}) \end{bmatrix} \quad \underline{X} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \quad \underline{B} = \begin{bmatrix} \lambda \\ \lambda A \\ \lambda B \\ \lambda AB \\ \lambda 11 \end{bmatrix}$$

Jede Spalte der Designmatrix \underline{X} (mit Ausnahme der ersten) vertritt einen unabhängigen Kontrast, der jeweils mit einem entsprechenden Parameter in \underline{B} korrespondiert. Jede logarithmierte Häufigkeit ist damit als lineare Funktion von Grundeffekten und dem Interaktionseffekt darstellbar. Die Bestimmung der Parameterwerte erfolgt durch eine einfache multiple Regression der Dummy-Variablen auf die Erwartungswerte einer Modellspezifikation über alle Subpopulationen (= Zellen der untersuchten Kontingenztafel) unter Verwendung z.B. der Effekt-Kodierung. Die resultierenden Regressionskoeffizienten werden auch als "zentrierte Effekte" bezeichnet.

Eine alternative Reparametrisierungsform ergibt sich bei der Verwendung einer 0/1-Kodierung. Dabei wird jeweils der Effekt einer ausgewählten Merkmalsausprägung (in der Regel die erste Kategorie) gleich Null gesetzt (daher der Ausdruck "auf eine Kategorie bezogene Effekte" oder engl.: "cornered effects"). Dabei entfallen etwa für das Unabhängigkeitsmodell zwei Spalten der Designmatrix \underline{X} , die danach wiederum einen vollen Spaltenrang aufweist. In unserem kleinen Beispiel einer 2 x 2-Kontingenztafel erhalten wir dann folgende Designmatrizen:

<u>UNABHÄNGIGKEITSMODELL</u>	<u>ZELLENHÄUFIGKEIT</u>	<u>ABHÄNGIGKEITSMODELL</u>
$\underline{X} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}$	$\begin{matrix} \hat{m}_{11} \\ \hat{m}_{21} \\ \hat{m}_{12} \\ \hat{m}_{22} \end{matrix}$	$\underline{X} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}$
EFFEKTE: v_0 v_A v_B		v_0 v_A v_B $v_{A \times B}$

wobei

	B = 1	B = 2
A = 1	\hat{m}_{11}	\hat{m}_{12}
A = 2	\hat{m}_{21}	\hat{m}_{22}

$\hat{m}_{ij} = n_{ij}$ im saturierten Modell

Gegenüberstellung der Reparametrisierungsbedingungen für das saturierte Modell:

Y	λ_A	λ_B	$\lambda_{A \times B}$	λ_0	v_A	v_B	$v_{A \times B}$	v_0
$\ln n_{11}$	1	1	1	1	0	0	0	1
$\ln n_{21}$	-1	1	-1	1	1	0	0	1
$\ln n_{12}$	1	-1	-1	1	0	1	0	1
$\ln n_{22}$	-1	-1	1	1	1	1	1	1

Daraus resultieren folgende Modellgleichungen für die einzelnen Zellen der 2 x 2 - Tabelle mit "centered effects":

UNABHÄNGIGKEITSMODELL:

$$\begin{aligned}\ln \hat{m}_{11} &= v_0 \\ \ln \hat{m}_{21} &= v_0 + v_A \\ \ln \hat{m}_{12} &= v_0 + v_B \\ \ln \hat{m}_{22} &= v_0 + v_A + v_B\end{aligned}$$

daher gilt:

$$\begin{aligned}v_A &= \ln \hat{m}_{21} - v_0 = \ln \hat{m}_{22} - \ln \hat{m}_{12} \\ v_B &= \ln \hat{m}_{12} - v_0 = \ln \hat{m}_{22} - \ln \hat{m}_{21}\end{aligned}$$

ABHÄNGIGKEITSMODELL:

$$\begin{aligned}\ln n_{11} &= v_0 \\ \ln n_{21} &= v_0 + v_A + v_{AB} \\ \ln n_{12} &= v_0 + v_B + v_{AB} \\ \ln n_{22} &= v_0 + v_A + v_B + v_{AB}\end{aligned}$$

daher gilt:

$$\begin{aligned}v_A &= \ln n_{21} - v_0 - v_{AB} = \ln n_{22} - \ln n_{12} \\ v_B &= \ln n_{12} - v_0 - v_{AB} = \ln n_{22} - \ln n_{21}\end{aligned}$$

($\ln \hat{m}_{ij}$) nsatz zu der Effekt-Kodierung können im Fall von "centered effects" die Parameterschätzwerte des Unabhängigkeitsmodells als Abweichungen der jeweils zweiten Ausprägung des Merkmals A bzw. B von der ersten Ausprägung zu interpretieren. Die Konstante v ist nicht mehr der Gesamtdurchschnitt der ($\ln \hat{m}_{ij}$)-Werte, sondern gerade $\ln \hat{m}_{11}$, also der Wert für die Ausprägungskombination A=1 und B=1.

Wir haben die Reparametrisierungsbedingungen bereits an dieser Stelle anhand der einfachen 2 x 2-Tabelle ausführlich gegenübergestellt, da in den EDV-Programmen zur log-linearen Tabellenanalyse (5) jeweils unterschiedliche Kodierungsformen für nominale/ordinale Merkmale vorgesehen sind.

Der Vorteil der log-linearen Modellbedingungen besteht darin, daß sie ohne Schwierigkeiten auf drei- oder höherdimensionale Kontingenztabellen zu erweitern ist. Der Schwerpunkt des Interesses bei der Zusammenhangsanalyse zweier nominaler Merkmale lag in der Frage, ob diese voneinander unabhängig sind oder nicht. Geht man nun über zur Analyse von mehr als zwei Variablen, so ist mit der erweiterten Variablenzahl

auch eine wesentliche Erweiterung möglicher Zusammenhänge möglich, die insgesamt die Zusammenhangsstruktur einer mehrdimensionalen Häufigkeitstabelle bilden.

Den nun folgenden abstrakteren Darlegungen des Goodman-Modells für höherdimensionale Tabellen wird zur Illustration der wichtigsten Grundgedanken folgendes exemplarisches Beispiel vorangestellt. Das aufbereitete Datenmaterial unserer Untersuchung enthält u.a. folgende Variablen für das Kaiserreich:

V: Beruf des Vaters

- (1) Besitz- und Bildungsbürger
- (2) Mittelstand alt und Mittelstand neu
- (3) Arbeiter und untere Beamte

B: Bildungsrang des Sohnes

- (1) Volksschule
- (2) Mittelschule bis Tertia
- (3) Sekunda und Prima

S: Herkunftsort

- (1) Minden
- (2) Duisburg

Alle Variablen werden nominal interpretiert. Das empirische Ausgangsmaterial unserer exemplarischen dreidimensionalen Tabellenanalyse stellt eine (2 x 3 x 3)-Tabelle dar, in der die zwei Merkmale BERUF DES VATERS und BILDUNG zusammen mit der Variablen STADT kreuzklassifiziert wurden; diese Tabelle läßt sich auch als Matrix der Zellenhäufigkeit n_{ijk}^{VBS} auffassen, die die vollständige Datenmatrix der drei Merkmale wiedergibt:

STADT	VATERBERUF	BILDUNG	n_{ijk}	STADT	VATERBERUF	BILDUNG	n_{ijk}
1	1	1	10	2	1	1	24
1	1	2	118	2	1	2	54
1	1	3	68	2	1	3	60
1	2	1	394	2	2	1	1969
1	2	2	891	2	2	1	633
1	2	3	251	2	2	3	247
1	3	1	1580	2	3	1	6578
1	3	2	449	2	3	2	270
1	3	3	21	2	3	3	19

STADT (S)	VAB (V)	BILDUNG		
		Volksschule	Mittelsch.-Tertia	Sek.-Prima
Minden	Besitzb. u. Bildungs.	10	118	68
	Mittelstand alt u. neu	394	89	251
	Arbeiter u. unt. Beamte	1580	449	21
Duisburg	Besitzb. u. Bildungs.	24	54	60
	Mittelstand alt u. neu	1969	633	247
	Arbeiter u. unt. Beamte	6578	270	19
		N = 13636		

Allgemein hat eine dreidimensionale Kontingenztafel folgende Gestalt:

C	B	A1	A2	...	AI	\sum	\sum
C1	B1	n_{111}	n_{211}	...	n_{I11}	$n_{\cdot 11}$	
	B2	n_{121}			\vdots	\vdots	$n_{\cdot \cdot 1}$
	\vdots				\vdots	\vdots	
	\vdots				\vdots	\vdots	
	Bj	n_{1j1}	...		n_{Ij1}	$n_{\cdot j1}$	
C2	B1	n_{112}	n_{212}	...	n_{I12}	$n_{\cdot 12}$	
	B2	n_{122}			\vdots	\vdots	$n_{\cdot \cdot 2}$
	\vdots				\vdots	\vdots	
	\vdots				\vdots	\vdots	
	Bj	n_{1j2}	...		n_{Ij2}	$n_{\cdot j2}$	
\vdots	\vdots	\vdots		n_{ijk}		\vdots	\vdots
CK	B1	n_{11K}	n_{21K}	...	n_{I1K}	$n_{\cdot 1K}$	
	B2	n_{12K}			\vdots	\vdots	$n_{\cdot \cdot K}$
	\vdots				\vdots	\vdots	
	\vdots				\vdots	\vdots	
	Bj	n_{1jK}	...		n_{IjK}	$n_{\cdot jK}$	
\sum		$n_{1\cdot \cdot}$...		$n_{I\cdot \cdot}$	$n_{\cdot \cdot \cdot} = N$	

$$n_{i\cdot \cdot} = \sum_{j=1}^J \sum_{k=1}^K n_{ijk}$$

$$n_{\cdot j\cdot} = \sum_{i=1}^I \sum_{k=1}^K n_{ijk}$$

$$n_{\cdot \cdot k} = \sum_{i=1}^I \sum_{j=1}^J n_{ijk}$$

$$n_{ij\cdot} = \sum_{k=1}^K n_{ijk}$$

$$n_{i\cdot k} = \sum_{j=1}^J n_{ijk}$$

$$n_{\cdot jk} = \sum_{i=1}^I n_{ijk}$$

$$n_{\cdot \cdot \cdot} = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K n_{ijk}$$

N = Stichprobenumfang

Analog zum zweidimensionalen log-linearen Modell können die logarithmierten zu erwartenden Häufigkeiten \hat{m}_{ijk}^{ABC} der dreidimensionalen Tabelle (A x B x C) als Summe von Parametern dargestellt werden. Wir wollen zunächst vom saturierten Modell ausgehen, das keine Restriktionen bzgl. bestimmter Parameter enthält:

$$A.14 \quad \ln \hat{m}_{ijk}^{ABC} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC}$$

Die Parameter erfüllen unter Annahme der Effekt-Kodierung die Bedingungen:

$$A.15 \quad \begin{aligned} \sum_i \lambda_i^A &= \sum_j \lambda_j^B = \sum_k \lambda_k^C = \sum_i \lambda_{ij}^{AB} = \sum_j \lambda_{ij}^{AB} = \dots = \\ \sum_i \lambda_{ijk}^{ABC} &= \sum_j \lambda_{ijk}^{ABC} = \sum_k \lambda_{ijk}^{ABC} = 0. \end{aligned}$$

Das saturierte Modell enthält ebensoviel frei variierende Parameter wie Zellen in der Kontingenztabelle; es ist damit den beobachteten Daten vollständig angepaßt (d.h. $n_{ijk} = \hat{m}_{ijk}^{ABC}$). Die Parameter in der Grundgleichung des saturierten Modells für den dreidimensionalen Fall sind wie folgt definiert:

$$\begin{aligned} \lambda &= -\frac{1}{IJK} \sum_{i,j,k} \ln \hat{m}_{ijk}^{ABC} \\ \lambda_i^A &= \frac{1}{JK} \sum_{j,k} \ln \hat{m}_{ijk}^{ABC} - \lambda \\ \lambda_j^B &= \frac{1}{IK} \sum_{i,k} \ln \hat{m}_{ijk}^{ABC} - \lambda \\ \lambda_k^C &= \frac{1}{IJ} \sum_{i,j} \ln \hat{m}_{ijk}^{ABC} - \lambda \\ \lambda_{ij}^{AB} &= \frac{1}{K} \sum_k \ln \hat{m}_{ijk}^{ABC} - \lambda_i^A - \lambda_j^B - \lambda \\ \lambda_{ik}^{AC} &= \frac{1}{J} \sum_j \ln \hat{m}_{ijk}^{ABC} - \lambda_i^A - \lambda_k^C - \lambda \\ \lambda_{jk}^{BC} &= \frac{1}{I} \sum_i \ln \hat{m}_{ijk}^{ABC} - \lambda_j^B - \lambda_k^C - \lambda \\ \lambda_{ijk}^{ABC} &= \ln \hat{m}_{ijk}^{ABC} - \lambda_i^A - \lambda_j^B - \lambda_k^C - \lambda_{ij}^{AB} - \lambda_{ik}^{AC} - \lambda_{jk}^{BC} - \lambda \end{aligned}$$

($\lambda =$

($\lambda_{ij}^{AB}, \lambda_{ik}^{AC}, \lambda_{jk}^{BC}$) Varianzanalyse en ($\lambda_i^A, \lambda_j^B, \lambda_k^C$) Modellp: (λ_{ijk}^{ABC}) einer Konstanten ($\lambda =$

Merkmalen sich verändert, wenn man von einer Kategorie des dritten Merkmals zu einer anderen übergeht. Aus der Grundgleichung A.14 ist ferner erkennbar, daß der log-lineare Modellansatz die analytische Trennung von Effekten der Randverteilung (Haupteffekte: λ_i^A, λ_j^B und λ_k^C) vom Einfluß der Beziehungen zwischen den Merkmalen (Zwei - Faktor - Interaktionen: $\lambda_{ij}^{AB}, \lambda_{ik}^{AC}$ und λ_{jk}^{BC}) erlaubt. Es wird vielfach nicht beachtet, daß einige der gebräuchlichen Maßzahlen zur Quantifizierung von Assoziationen in den Kontingenztabellen nicht nur von der Stärke der Beziehungen abhängen, sondern auch von den Randverteilungen der betrachteten Merkmale beeinflusst werden können.

Analog zum zweidimensionalen Fall läßt sich das saturierte Modell z.B. für eine (2 x 2 x 2)-Tabelle in der Designmatrixdarstellung überführen.

$$\begin{bmatrix} \ln \hat{m}_{111} \\ \ln \hat{m}_{211} \\ \ln \hat{m}_{121} \\ \ln \hat{m}_{221} \\ \ln \hat{m}_{112} \\ \ln \hat{m}_{212} \\ \ln \hat{m}_{122} \\ \ln \hat{m}_{222} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 \\ 1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & -1 & 1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} \lambda \\ \lambda^A_1 \\ \lambda^B_1 \\ \lambda^C_1 \\ \lambda^{AB}_{11} \\ \lambda^{AC}_{11} \\ \lambda^{BC}_{11} \\ \lambda^{ABC}_{111} \end{bmatrix}$$

Die Parameter des saturierten Modells sind wie folgt zu interpretieren:

Das Gesamtmittel (λ) der logarithmierten erwarteten Häufigkeiten ergibt sich als Mittelwert über alle $\ln \hat{m}_{ijk}$.

Die Haupteffekte sind Funktionen, die die Abweichungen vom Gesamtmittel λ für alle Zellen angeben, die den Index i bzw. j oder k enthalten. So kennzeichnet z.B. λ^A_1 die durch das erste Merkmal in der Ausprägung i bedingte Abweichung der logarithmierten erwarteten Häufigkeiten vom Gesamtmittel.

Die Interaktionseffekte 1.Ordnung (oder Zwei-Faktor-Effekte) kennzeichnen den Beitrag, den die gemeinsame Betrachtung jeweils zweier Merkmale über die bereits dargestellten Effekte hinaus liefert. Die Zusammenhänge zwischen A und B, B und C sowie zwischen A und C werden durch diese Interaktionseffekte bestimmt.

Der Interaktionseffekt 2.Ordnung (Drei-Faktor-Interaktion) besagt, daß der Zusammenhang zwischen je zwei Merkmalen von den Ausprägungen des verbleibenden Merkmals abhängt. Der Zusammenhang zwischen Merkmal A und B wird also durch die

Interaktionsparameter λ_{AB} und durch die Drei-Faktor-Interaktion λ_{ABC} , die die Modifikation des Zusammenhangs durch die Variable C verkörpert, bestimmt. Der Zusammenhangsstruktur kann so z.B. folgende Form annehmen:



Interessant für unsere Analyse der Tabelle (VAB x STADT x BILDUNG) sind nun Modelle, die nicht sämtliche möglichen Effekte enthalten. Die Strategie im Rahmen log-linearer Tabellenanalyse ist gerade, ein Modell zu suchen, das so einfach wie möglich ist, d.h. wenige höherdimensionale Effekte enthält, und zugleich den empirischen Daten hinreichend entspricht. Alle Modelle mit signifikanten Abweichungen sollen daher aufgrund ihrer mangelnden Anpassung an den empirischen Befund zurückgewiesen werden. Bei der Wahl des Signifikanzniveaus sollte wiederum berücksichtigt werden, daß man einen Anpassungstest durchführt und explizit nicht an signifikanten Ergebnissen interessiert ist. Zu niedrige Signifikanzniveaus (z.B. der übliche 5%-Wert) würden Modelle mit hohen Abweichungen nicht von der Modellsuche ausschließen. Einige Autoren schlagen daher vor, ein Signifikanzniveau (z.B. 25%) zu wählen, das es erlaubt, schon kleine Abweichungen vom wahren Modell festzuhalten (z.B. Forthofer/Lehnen 1981). Als Faustregel kann gelten, daß man Modelle als passend akzeptiert, bei denen die Abweichungen kleiner oder gleich der Anzahl der Freiheitsgrade ist. Die Vorgehensweise sei anhand einer dreidimensionalen Kontingenztafel $X \times Y \times Z$ illustriert. Bei der Spezifikation nichtsaturierter Modelle sind zunächst folgende Punkte von Bedeutung:

1. Zusammenhänge zwischen jeweils zwei Merkmalen (Interaktionen oder Wechselwirkungen) werden jeweils paarweise mit dem jeweiligen Anfangsbuchstaben der beteiligten Merkmale bezeichnet (sogenannte angepaßte Randverteilungen).

Beispiel: (XY, YZ, XZ)

Die angepaßten Randverteilungen XY, YZ und XZ entsprechen in der log-linearen Tabellenanalyse jenen Parametern, für die der korrespondierende Effekt ungleich Null angenommen wird. In dem Modell-Beispiel würde somit jeweils von einem paarweisen Zusammenhang zwischen den Merkmalen X, Y und Z ausgegangen (Modellhypothese).

2. Bei der Anwendung des log-linearen Verfahrens kommt es darauf an, ein möglichst sparsames Modell zu finden, das den Daten noch hinreichend angepaßt ist, d.h. die Abweichungen zwischen den beobachteten Häufigkeiten (n_{ijk}) und modellimplizierten Häufigkeiten (\hat{m}_{ijk}) sollten nicht zu groß sein. Hat man somit ein Modell mit den erwarteten Häufigkeiten berechnet, so ist zu prüfen, ob

diese von den empirischen Häufigkeiten gemessen am Wert der χ^2_L - Statistik signifikant abweichen:

$$\chi^2_L = G^2 = 2 \sum_i \sum_j \sum_k n_{ijk} \ln (n_{ijk} / \hat{m}_{ijk})$$

3. Wie beschränken uns im folgenden auf sogenannte hierarchische Modelle, bei denen zu jedem Effekt einer bestimmten Ordnung auch alle durch Aufspaltung gewonnenen Effekte niedriger Ordnung im Modell spezifiziert sind. Ist z.B. der Interaktionseffekt zwischen X und Y, der mit (XY) bezeichnet wird, in der Modellhypothese enthalten, so müssen auch die Haupteffekte von X (also (X)) und von Y (also (Y)) berücksichtigt werden.

Das saturierte Modell (XYZ) enthält somit alle Effekte niedriger Ordnung:

- XY, XZ, ZY
- X, Y, Z

(Daher läßt sich auch ein log-lineares Modell durch die maximalen Effekte zur Abkürzung der Modellhypothese verwenden, s. Punkt 1).

4. Verschiedene hierarchische Modelle, die sich durch einen unterschiedlichen Komplexitätsgrad bzgl. der Parameterzahl auszeichnen, lassen sich zu einer geschachtelten Modellsequenz ("nested sequence") anordnen, in der sukzessive Parameter höherer Ordnung spezifiziert werden.

Die Tabelle auf der folgenden Seite illustriert mögliche Hypothesen für eine dreidimensionale Tabelle, wobei das Schaubild den Hierarchiecharakter in Mengendarstellung veranschaulicht:

- Wir können die Modellhierarchie wie folgt charakterisieren:
 - + Der Parameterset S (XY) (XZ) (YZ) ist eine Teilmenge des saturierten Modells (XYZ);
 - + Der Parameterset S (XY) (XZ) ist wiederum eine Teilmenge des Modells (XY) (XZ) (YZ);
 - + Der Parameterset S (XY,Z) ist eine Teilmenge von S (XY) (XZ) usw.
- Vernachlässigt man von einem bestimmten Modell einige Parameter, so ergibt sich ein spezielles Submodell mit der Eigenschaft, daß seine Verträglichkeit mit dem empirischen Befund (gemessen an den χ^2_L -Wert) auch die Gültigkeit des ursprünglichen Modells impliziert.

Wenn man z.B. das Submodell:

M 1: (X) (Y) (Z)

von

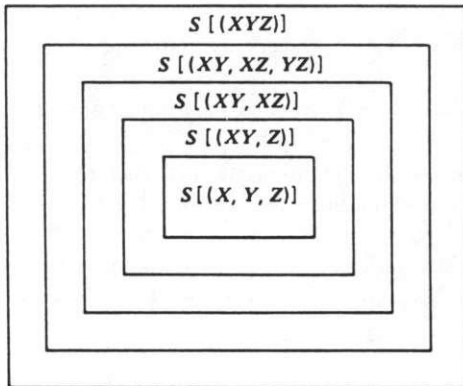
M 2: (XY) (Z)

betrachtet, so ist, wenn M 1 von den beobachteten Daten her nicht zurückgewiesen werden kann, natürlich auch M 2 in diesem Sinne ein gültiges Modell,

Ausgewählte Hypothesen in einer dreidimensionalen Tabelle X x Y x Z:

MODELLSEQUENZ	MODELL-SYMBOL
$\log \hat{m}_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z$	(X, Y, Z)
$\log \hat{m}_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY}$	(XY, Z)
$\log \hat{m}_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{jk}^{YZ}$	(XY, YZ)
$\log \hat{m}_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{jk}^{YZ} + \lambda_{ik}^{XZ}$	(XY, YZ, XZ)
$\log \hat{m}_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{jk}^{YZ} + \lambda_{ik}^{XZ} + \lambda_{ijk}^{XYZ}$	(XYZ)

Hierarchische Modelle unterschiedlichen Komplexitätsgrades:



Es gilt somit:

$$S[(X, Y, Z)] \subset S[(XY, Z)] \subset S[(XY, YZ)] \\ \subset S[(XY, XZ, YZ)] \subset S[(XYZ)]$$

denn der hinzutretende Parameter λ_{ij}^{XY} führt in jedem Fall noch zu einer weiteren Verbesserung der Modellanpassung.

- Aus einer Sequenz von geschachtelten Modellen läßt sich die χ^2_L -Anpassungsstatistik in Komponenten zerlegen ("partitioning χ^2_L "). Als Beispiel betrachten wir folgende Modellsequenz:

H1: (XY) (YZ) (XZ)

H2: (XY) (YZ)

H3: (XY) (Z)

H4: (X) (Y) (Z)

$$\begin{aligned}
 G^2[(X, Y, Z)] &= G^2[(XY, YZ, XZ)] \\
 &\quad + \{G^2[(XY, YZ)] - G^2[(XY, YZ, XZ)]\} \\
 &\quad + \{G^2[(XY, Z)] - G^2[(XY, YZ)]\} \\
 &\quad + \{G^2[(X, Y, Z)] - G^2[(XY, Z)]\} \\
 &= G^2[(XY, YZ, XZ)] \\
 &\quad + G^2[(XY, YZ) | (XY, YZ, XZ)] \\
 &\quad + G^2[(XY, Z) | (XY, YZ)] \\
 &\quad + G^2[(X, Y, Z) | (XY, Z)]
 \end{aligned}$$

Für die Praxis der Modell-Suche folgt aus der CHI^2_L -Zerlegung z.B. folgende Vorgehensweise:

1. Schritt: Überprüfen, ob das Modell $(XY)(YZ)(XZ)$ dem empirischen Befund angemessen ist.
2. Schritt: Wenn ja, unter Vernachlässigung der Interaktion (XZ) überprüfen, ob eine nennenswerte Verschlechterung der Modellanpassung resultiert.
3. Schritt: Wird das reduzierte Modell nicht von den beobachteten Daten zurückgewiesen, wird überprüft, inwieweit der Term (YZ) vernachlässigt werden kann usw..

Sukzessive gelangt man über eine geschachtelte Modellsequenz zu einem immer sparsameren Modell, wobei die additive Zerlegung der CHI^2_L -Statistik eben nur für "nested models" gilt (ein jeweils sparsameres Modell umfaßt eine Parameterteilmenge aus einem in der Hierarchie vorausgehenden Modellansatz). Da die Zerlegung auch für die Freiheitsgrade gilt, läßt sich die Differenz zweier CHI^2_L -Werte aus einem Modellvergleich selbst auch auf Signifikanz prüfen.

Beispiel: $G^2(H_2) - G^2(H_1)$ ist
 χ^2 -verteilt mit
 d.f. Diff. = d.f. (H_2) - d.f. (H_1)
 Freiheitsgraden.

Ist die Differenz signifikant, muß das Modell H_1 beibehalten werden, d.h. die Interaktion (XZ) darf nicht gleich Null gesetzt werden.

In dem gewählten Beispiel sind selbstverständlich auch neben der hier behandelten Modellsequenz noch weitere geschachtelte Modellsequenzen überprüfbar. Wir wollen jedoch zunächst einmal die (abstrakte) Darstellung hier abbrechen und uns dem bereits vorgestellten Beispiel näher widmen.

Die folgende Tabelle enthält für 17 nicht-saturierte Modelle die χ^2 -Werte bzgl. der Abweichungen der Beobachtungswerte von den modellimplizierten Werten, die Freiheitsgrade (d.f.) und die korrespondierenden Signifikanzniveaus ("probability level"). Bevor wir auf die einzelnen Modellimplikationen und Testergebnisse näher eingehen, sei nochmals eine wichtige Eigenschaft log-linearer Modellbildung hervorgehoben. Es besteht eine eindeutige Zuordnung zwischen jenen Effekten, die in einem Modell berücksichtigt worden sind und jenen Randverteilungen, die durch das Modell exakt wiedergegeben werden (die sogenannten angepaßten oder suffizienten Randvertei-

MODELL	DF	LIKELIHOOD-RATIO CHISQ	PROB.
B.	15	17815.32	0.0
V.	15	20068.96	0.0
S.	16	27163.37	0.0
B,V.	13	7913.22	0.0
V,S.	14	17261.27	0.0
S,B.	14	15007.63	0.0
B,V,S.	12	5105.53	0.0
BV.	9	4547.06	0.0
BS.	12	13284.90	0.0
VS.	12	16917.21	0.0
B,VS.	10	4761.47	0.0
V,BS.	10	3382.80	0.0
S,BV.	8	1739.37	0.0
BV,BS.	6	16.64	0.0107
BS,VS.	8	3038.74	0.0
VS,BV.	6	1395.31	0.0
BV,BS,VS.	4	2.06	0.7072
B = Bildung; V = Beruf des Vaters; S = Stadt.			

lungen). Für das Unabhängigkeitsmodell unserer dreidimensionalen Tabelle, das lediglich die Effekte λ_i^B , λ_j^V und λ_k^S enthält (symbolisiert durch $B \otimes V \otimes S$), sind die Randverteilungen der Merkmale BILDUNG (B), STADT (S) und VAB (V) exakt enthalten. Wegen dieser eindeutigen Zuordnung kann eine Modellhypothese statt durch die enthaltenen Effekte auch durch Angabe der Randverteilungen, die im Modell durch Anpassung den empirischen entsprechen, beschrieben werden. Für die möglichen Hypothesen in unserem dreidimensionalen exemplarischen Beispiel erhalten wir somit eine Modellsequenz, die eine hierarchische Struktur beinhaltet (s. S. 36).

Unabhängigkeitsmodelle	log-lineare Modelle	Angepaßte Randverteilungen
<u>völlige Unabhängigkeit:</u> $\hat{m}_{ijk} = N(p_{i..} p_{.j.} p_{..k})$	$\ln \hat{m}_{ijk} = \lambda + \lambda_i^B + \lambda_j^V + \lambda_k^S$	(B) (V) (S)
<u>partielle Unabhängigkeit:</u> $\hat{m}_{ijk} = N(p_{.jk} p_{i..})$	$\ln \hat{m}_{ijk} = \lambda + \lambda_i^B + \lambda_j^V + \lambda_k^S + \lambda_{jk}^{VS}$	(VS) (B)
$\hat{m}_{ijk} = N(p_{i.k} p_{.j.})$	$\ln \hat{m}_{ijk} = \lambda + \lambda_i^B + \lambda_j^V + \lambda_k^S + \lambda_{ik}^{BS}$	(BS) (V)
$\hat{m}_{ijk} = N(p_{ij.} p_{..k})$	$\ln \hat{m}_{ijk} = \lambda + \lambda_i^B + \lambda_j^V + \lambda_k^S + \lambda_{ij}^{BV}$	(BV) (S)
<u>bedingte Unabhängigkeit:</u> $\hat{m}_{ijk} = N(p_{i.k} p_{.jk} / p_{..k})$	$\ln \hat{m}_{ijk} = \lambda + \lambda_i^B + \lambda_j^V + \lambda_k^S + \lambda_{ik}^{BS} + \lambda_{jk}^{VS}$	(BS) (VS)
$\hat{m}_{ijk} = N(p_{ij.} p_{.jk} / p_{.j.})$	$\ln \hat{m}_{ijk} = \lambda + \lambda_i^B + \lambda_j^V + \lambda_k^S + \lambda_{ij}^{BV} + \lambda_{jk}^{VS}$	(BV) (VS)
$\hat{m}_{ijk} = N(p_{ij.} p_{i.k} / p_{..i})$	$\ln \hat{m}_{ijk} = \lambda + \lambda_i^B + \lambda_j^V + \lambda_k^S + \lambda_{ij}^{BV} + \lambda_{ik}^{BS}$	(BV) (BS)
<u>keine 3-Faktoren-Interaktion:</u> $\hat{m}_{ijk} = N(p_{i.k} p_{.jk} p_{..k} / p_{..k})$	$\ln \hat{m}_{ijk} = \lambda + \lambda_i^B + \lambda_j^V + \lambda_k^S + \lambda_{ij}^{BV} + \lambda_{ik}^{BS} + \lambda_{jk}^{VS}$	(BV) (BS) (VS)

Dieses hierarchische Prinzip ist unmittelbar einleuchtend, wenn man die eindeutige Zuordnung von Effekten und angepaßten Randverteilungen bedenkt: aus der Randverteilung der drei Variablen B S V kann man alle darin enthaltenen Randverteilungen berechnen; aus diesem Grunde impliziert das Vorhandensein des Effektes λ_{ijk}^{BVS} auch das Vorhandensein folgender Effekte: $\lambda_i^B, \lambda_j^V, \lambda_k^S, \lambda_{ij}^{BV}, \lambda_{jk}^{VS}, \lambda_{ik}^{BS}$.

Deshalb reicht es zur Festlegung eines bestimmten Modells aus, die höchsten Wechselwirkungen in Form der angepaßten Randverteilungen anzugeben, die noch berücksichtigt werden sollen. Ein Modell, bei dem die Randverteilungen (BV) (SB) abgepaßt sind, enthält dann z.B. die Effekt-Parameter $\lambda_i^B, \lambda_j^V, \lambda_k^S, \lambda_{ij}^{BV}$ und λ_{ki}^{SB} .

Was besagen die einzelnen Modellannahmen in unserem Beispiel im einzelnen (7)? In dem Modell der vollständigen Unabhängigkeit (B S V) werden alle Interaktionsterme vernachlässigt, d.h. gleich Null angenommen:

H.: es gibt keine paarweisen Zusammenhänge zwischen den Merkmalen.

Die verschiedenen Herkunftsorte, Bildungsabschlüsse und Väterberufe kommen zwar unterschiedlich häufig vor, aber weder hat der Beruf des Vaters einen Einfluß auf den erreichten Bildungsrang, noch haben die unterschiedlichen Herkunftsorte einen Einfluß auf die Bildungschancen. Zwischen dem Beruf des Vaters und dem Herkunftsort besteht ebenfalls kein Zusammenhang. Die χ^2 -Testgröße zeigt, daß diese (unrealistische) Annahme verworfen werden muß (χ^2 : 5105.5; D.F.: 12; $p = 0.0$). Dieses sehr restriktive Modell bildet nun den Ausgangspunkt, von dem aus untersucht wird, in welchem Ausmaß die Divergenz zwischen beobachteten und erwarteten Häufigkeiten vermindert wird, wenn wir sukzessive Interaktionsterme in dieses 'Basismodell' aufnehmen. Wir setzen die Abweichung von den beobachteten Daten gleich 100% und untersuchen, welcher Anteil dieser Divergenz bei der Formulierung realistischer Modelle erklärt bzw. reduziert werden kann (8).

Betrachten wir nun die Modelle der partiellen Unabhängigkeit, in denen jeweils ein Interaktionsterm für den Zusammenhang zwischen zwei Merkmalen gegenüber dem Basismodell hinzutritt:

Unabhängigkeitshypothese:	Erklärte Devianz gegenüber dem Basismodell (S) (V) (B)
B ⊗ VS → (B) (VS)	6,7%
V ⊗ BS → (V) (BS)	33,7%
S ⊗ BV → (S) (BV)	65,9%

" ⊗ " symbolisiert: ist unabhängig von (9).

Keines der Modelle weist eine ausreichende Modellanpassung auf. Wir erweitern daher die Modelle der partiellen Unabhängigkeit um einen weiteren Interaktionsterm. Mit dieser Erweiterung erhalten wir die folgenden Modelle der bedingten Unabhängigkeit (" | " symbolisiert: unter Bedingung):

Unabhängigkeitshypothese:	Erklärte Devianz gegenüber dem Basismodell (S) (V) (B)
V ⊗ S B → (VB) (SB)	99,6%
B ⊗ V S → (BS) (VS)	40,5%
S ⊗ B V → (SV) (BV)	72,7%

Die erste Hypothese konditionaler Unabhängigkeit

H_0 : Die Variable BERUF DES VATERS ist unabhängig von der Variablen STADT für alle Ausprägungen der Variablen BILDUNG

ist zwar nach dem Signifikanzkriterium abzulehnen ($p < 0,25$), die erklärte Devianz gegenüber dem Basismodell ist aber mit einem Wert von 99,6% so hoch, daß wir dieses Modell dennoch akzeptieren können. Wir wollen abschließend das Modell mit den Interaktionseffekten λ_{ji}^{VB} und λ_{ki}^{SB} interpretieren.

Als wichtiges Ergebnis kann zunächst festgehalten werden, daß es wie erwartet einen Zusammenhang zwischen BILDUNG und BERUF DES VATERS gibt. Interessant ist auch die zu berücksichtigende Interaktion zwischen BILDUNG und STADT: Der erreichte Bildungsabschluß hängt auch von der Herkunftsstadt ab. Der Term "konditionale Unabhängigkeit zwischen den Variablen VAB und STADT besagt, daß VAB für jede Ausprägung der Variablen BILDUNG ($B_i = 1,2,3$) unabhängig von STADT ist. Wenn jedoch zwischen VAB und BILDUNG als auch zwischen STADT und BILDUNG ein Zusammenhang besteht, könnte sich bei ausschließlicher Betrachtung der Variablen VAB und STADT in einer aggregierten zweidimensionalen Kontingenztabelle durchaus ein Scheinzusammenhang ergeben, der nur durch die Vernachlässigung der Variablen BILDUNG entstanden ist. Für diesen Modelltyp gilt bzgl. der ersten Variante ferner:

- Die Assoziation zwischen VAB und BILDUNG hängt nicht von den Ausprägungen der Variablen STADT ab.
- Die Assoziation zwischen STADT und BILDUNG hängt nicht von den Ausprägungen der Variablen VAB ab.

Wäre das Modell (VB) (SB) nicht mit dem empirischen Befund verträglich gewesen, so hätten wir zu dem Modell ohne Drei-Faktor-Interaktion

(BV) (BS) (VS)

übergehen müssen (Erweiterung des Modells um die Assoziation zwischen STADT und VAB). In diesem Modell ist somit eine Assoziation je zweier Variablen zulässig. Eine Vereinfachung gegenüber dem saturierten Modell ergibt sich aber insofern, als die Art des Zusammenhangs (d.h. die Form und die Stärke des Zusammenwirkens jeweils zweier Merkmale) jeweils von dem dritten Merkmal unabhängig ist:

$B \leftrightarrow V$ ist gleichartig für STADT = 1,2 ,

$B \leftrightarrow S$ ist gleichartig für VAB = 1,2,3 ,

$V \leftrightarrow S$ ist gleichartig für BILDUNG = 1,2,3 .

Formal ausgedrückt: Die bedingte Assoziation zwischen zwei Merkmalen hängt nicht von der verbleibenden Variablen ab. Diese Modelleigenschaft wird in unserem dreidimensionalen Beispiel mit der Annahme des saturierten Modells aufgegeben.

Was die relative Bedeutung der zwei Interaktionseffekte betrifft, ergibt sich aus den Ergebnissen ein sehr eindeutiges Bild. Wird das Modell der Unabhängigkeit um den Interaktionseffekt zwischen BILDUNG und BERUF DES VATERS erweitert, erreichen wir bereits eine erklärte Devianz von gerundet 66%. Knapp mehr als die Hälfte an

empirischer Variation in der zugrundeliegenden Tabelle bestimmt demnach der Zusammenhang zwischen sozialer Herkunft und erreichten Bildungsrang. Etwa halb so stark ist der Zusammenhang zwischen BILDUNG und STADT. Der zusätzliche Anteil erklärter Devianz folgt aus

$$\frac{\chi^2_L \text{ (BV) (S)} - \chi^2_L \text{ (BV) (BS)}}{\chi^2_L \text{ (V) (S) (B)}} = \frac{1739 - 16,6}{5105} = 33,7\%$$

Auf die Berechnung und Diskussion der Lambda-Parameter in unserem bestätigten Modell ((BV) (BS)) wollen wir hier zunächst verzichten, da wir in dem folgenden Abschnitt ausführlich auf deren Interpretation im Rahmen einer Abhängigkeitsanalyse eingehen.

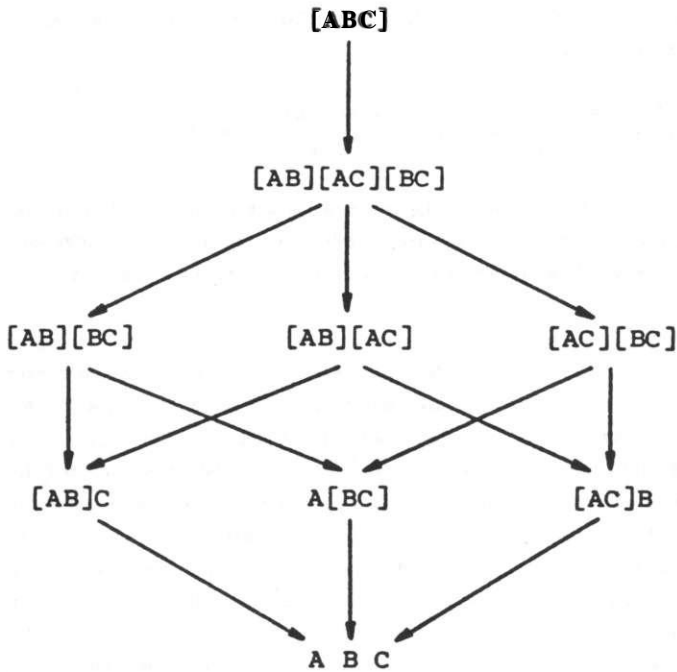
Abschließend wollen wir noch zusammenfassend die Familie hierarchisch geordneter Modelle allgemein für eine dreidimensionale Tabelle (A x B x C) darstellen, wobei wir wiederum jeweils die maximalen Effekte zur Abkürzung der Modelle verwenden. In der Abbildung auf der folgenden Seite ist ein Baumdiagramm wiedergegeben, daß die möglichen Modellsequenzen im dreidimensionalen Fall illustriert. Je zwei Modelle, die durch einen Pfeil oder eine Folge von Pfeilen miteinander verbunden sind, stehen in hierarchischer Ordnung zueinander, wobei die Pfeile in Richtung des sparsameren Modells weisen. Für die praktische Arbeit ist es außerordentlich nützlich, eine mit dem Diagramm korrespondierende Arbeitstabelle anzulegen, die alle wesentlichen χ^2_L -Anpassungssstatistiken systematisiert (s.d. Tabelle unterhalb des Diagramms):

- χ^2_L -Werte und p-Werte für das jeweils überprüfte Modell;
- χ^2_L -Differenzen zwischen jeweils zwei Modellen innerhalb einer Sequenz.

Für die Suche nach einem möglichst sparsamen Modell, das den beobachteten Daten aber noch hinreichend angepaßt sein muß, lassen sich folgende einfache Regeln aufstellen:

1. Man beginnt bei der multivariaten Tabellenanalyse mit einem Maximalmodell (saturiertes Modell, das den empirischen Befund perfekt reproduziert) und versucht es sukzessive zu vereinfachen, ohne daß die Abweichungen zwischen beobachteten und modellimplizierten Häufigkeiten zu groß werden. Der χ^2_L -Anpassungstest sollte p-Werte größer 0.25 (Signifikanzniveau) aufweisen.
2. Signifikante χ^2_L -Differenzen beim Übergang zu einem sparsameren Modell deuten darauf hin, daß der jeweils vernachlässigte Interaktionseffekt eine signifikante Erhöhung der Devianz zur Folge hat (das Ursprungsmodell müßte somit beibehalten werden).

Baundiagramm möglicher Sequenzen von (geschachtelten) Modellen im Rahmen einer dreidimensionalen Tabellenanalyse:



Arbeitstabelle für die Modellsequenzen:

	MODELL	D.F.	χ^2_L	P	NÄCHSTE VERZWEIGUNG
1	[ABC]				2< $\xleftarrow{\chi^2_L\text{-Differenz}}$
2	[AB][BC][AC]				3< >
					4< >
					5< >
3	[AB][AC]				6< >
					7< >
4	[AB][BC]				6< >
					8< >
5	[AC][BC]				7< >
					8< >
6	[AB]C				9< >
7	[AC]B				9< >
8	A[BC]				9< >
9	A B C				

3. Unabhängig vom Ausgang der Modelltests sollte man die Anpassung der Modelle an die beobachteten Häufigkeiten zusätzlich durch deskriptive Maße beschreiben (Verbesserung der Modellanpassung relativ zum Basismodell der vollständigen Unabhängigkeit, ausgedrückt im % Anteil der Modellverbesserung).

Kontingenztabellen mit vier und mehr Dimensionen lassen sich schließlich völlig analog zum dreidimensionalen Fall im Rahmen der log-linearen Modellbildung behandeln (d.h. in den Kategorien "Hypothesentyp", Abkürzung der jeweiligen Modellannahme in Form "angepaßter Randverteilungen" etc.), wobei folgende (allgemeine) Prinzipien der Modellselektion an Bedeutung gewinnen (s. die vergleichende Diskussion in Hamerle/Tutz 1984, S. 532ff.):

Unter Einbeziehung der Methoden zur Beurteilung der Modellanpassung (Anpassungstest, prozentuale Verbesserung der Anpassung relativ zum Basismodell, Signifikanz einzelner Assoziationsterme anhand bestimmter CHI - Differenzen) sollte ein geeignetes Modell:

- (1) sparsam (einfach) sein, d.h. so wenig wie möglich Parameter enthalten, um eine ausreichende Anpassung (Modell-Fit) zu gewährleisten, so daß ein Test auf Güte der Modellanpassung das entsprechende Modell nicht ablehnt,
- (2) alle signifikanten Effekte enthalten, d.h. fügt man zu einem weniger komplexen Modell einen zusätzlichen Modellparameter ein, so sollte die durch diesen Effekt verursachte Veränderung der Modellanpassung nicht signifikant sein. Eliminiert man jedoch einen fraglichen Parameter aus dem Modell, so sollte das zu einer signifikanten Modellverschlechterung führen,
- (3) eine hinreichend gute Modellanpassung besitzen. Üblicherweise wählt man zur Beurteilung der Modellanpassung die ML-Anpassungsstatistik bzw. den entsprechenden p-Wert des Modells. Diese Werte sind jedoch stark beeinflusst vom Stichprobenumfang N und verschlechtern sich bei gleicher Anpassungsgüte mit wachsendem N. Daher wird bei sehr großem Stichprobenumfängen fast jedes Modell durch ein Testverfahren abgelehnt, auch wenn es eine gute Anpassung relativ zu einem Basismodell hat. In diesen Fällen sollten statt einer Testentscheidung auch untere Entscheidungskriterien gewählt werden, z.B. die prozentuale Modellverbesserung relativ zum Basis- oder Grundmodell.

In Abhängigkeit von diesen unterschiedlichen Zielkriterien gibt es verschiedene, in der Literatur ausführlich diskutierte Strategien zur Modell- und Variablenselektion (vgl. den Überblick in Hamerle/Tutz a.a.O.):

- (1) Variablenselektionsverfahren sollen aus einer relativ großen Variablenmenge die zur späteren Modellbildung relevanten Variablen herausfiltern bzw. irrelevante Variablen ausschließen.

- (2) Modellbildungsverfahren hingegen kommt die Aufgabe zu, aus einer fest vorgegebenen Variablenauswahl, endgültig (d.h. abschließend) ein geeignetes Modell auszuwählen.

Die bekannten Strategien zur Modell- bzw. Variablenselektion verlaufen nahezu alle nach dem selben Prinzip: Ausgehend von einem Basis- oder geeigneten Ausgangsmodell werden entweder sukzessive zusätzliche Parameter in das Modell aufgenommen, bis die Modellanpassung hinreichend gut genug ist ("forward selection") oder "überflüssige" Parameter aus dem Modell entfernt, solange die Modellanpassung nicht entscheidend verschlechtert wird, d.h. der Test zur Modellablehnung führen würde ("backward elimination"). Die Unterschiede zwischen den einzelnen Verfahren bestehen daher fast immer

- in der Richtung der Vorgehensweise (vorwärts/rückwärts),
- in der Festlegung des Ausgangsmodells,
- in der Wahl einer Test-Statistik zur Bestimmung der kritischen Parameter.

Besondere Beachtung verdienen schließlich noch unbesetzte Zellen. Der log-lineare Ansatz geht davon aus, daß jede Merkmalskombination eine positive Auftretenswahrscheinlichkeit besitzt. Wenn diese Bedingung erfüllt ist, sind in einer genügend großen Stichprobe keine leeren Zellen zu erwarten. Da in der Praxis jedoch der Stichprobenumfang beschränkt ist, treten insbesondere bei höherdimensionalen Tabellen leere Zellen ("sampling" oder "random zeros") auf. Während für eine leere Zelle die geschätzte (d.h. modellimplizierte) Zellenhäufigkeit m nicht zwangsläufig gleich Null ist, kann dieser Fall jedoch bei mehreren Zellen gleich Null auftreten. Ergibt ein Modell eine erwartete Häufigkeit von Null, können bestimmte LAMBDA-Parameter nicht mehr berechnet werden, da sich z.B. in \hat{m}_{ijk} für $\hat{m}_{ijk} = 0$ nicht mehr bilden läßt. Um diese Schwierigkeit zu umgehen, werden meist ad hoc - Prozeduren zur "Glättung" der Kontingenztabelle verwendet. So wird z.B. empfohlen, vor der Analyse im Fall leerer Zellen zu jeder Zelle 0.5 zu addieren oder leere Zellen durch $1/R$ (R = Gesamtzahl der Zellen) vorzubersetzen. Derartige Prozeduren sind praktikabel und notwendig, wenn man an Schätzwerten für die LAMBDA-Parameter interessiert ist (10). Vom statistischen Standpunkt aus betrachtet sind diese Prozeduren allerdings wenig befriedigend, denn der χ^2_{L} -Anpassungstest ist an bestimmte Voraussetzungen gebunden. Diese bestehen gerade darin, daß die Zellenhäufigkeiten (bis auf wenige Ausnahmen) gewisse Mindestgrößen nicht unterschreiten sollten. Als Daumenregel wird meist eine Mindestbesetzung von 5 genannt. Diese Bedingung setzt der Variablenzahl einer Kontingenztabelle oftmals enge Grenzen. Die genannten Einschränkungen ergeben sich jedoch nur, wenn von der Prüfstatistik (p-Wert) Gebrauch gemacht wird.

B. REGRESSIONSMODELLE FÜR DISKRETE ABHÄNGIGE MERKMALE:
INDIVIDUALDATENANALYSE IM VERGLEICH ZUR MEHRDIMENSIONALEN
KONTINGENZTABELLENANALYSE

Wir wollen in diesem Abschnitt zunächst die symmetrische Fragestellung zugunsten einer Abhängigkeitsanalyse aufgeben: Es sollen spezielle Modelle vorgestellt werden, in denen eine nominal oder ordinal skalierte Variable explizit als abhängige Größe in die Untersuchung eingeht. Konkret handelt es sich dabei um die Anwendung der üblicherweise nur bei der Analyse metrischer abhängiger Merkmale verwendeten einfachen und multiplen Regressionsanalyse auf diskrete Zielvariablen. Im folgenden werden die dabei auftretenden statistischen Probleme und die sich daraus ergebenden Besonderheiten bei der Ergebnisinterpretation erörtert.

Zur Vereinfachung der Ausgangssituation betrachten wir im folgenden lediglich eine dichotome Zielvariable (d.h. ein diskretes Merkmal mit zwei Ausprägungen) auf der Ebene von Individualdaten. Wir gehen somit von n nicht-aggregierten Untersuchungseinheiten einer Stichprobe aus. Die Ausprägungen einer dichotomen Zielvariablen (in der Regel mit "1" und "2" kodiert) lassen sich analog zu nicht-metrischen unabhängigen Merkmalen in eine 0/1-kodierte Dummy-Variable transformieren:

<u>Ursprüngliche Variable:</u>		<u>Dummy-Variable:/</u>
Wert 2	→	Wert 1
Wert 1	→	Wert 0

Eine solche Variable läßt sich z.B. künstlich festlegen, in dem immer dann, wenn bei einer Untersuchungseinheit ein bestimmtes Ereignis beobachtet wird, die abhängige Variable (künftig mit Y symbolisiert) den Wert "1" erhält, während wir ihr den Wert "0" zuweisen, wenn das betreffende Ereignis nicht beobachtet wird. Typische Fragestellungen in der empirischen Sozialforschung wären z.B.:

- die Analyse des generativen Verhaltens:
0 = kein Kind, 1 = ein Kind oder mehr als ein Kind;
- die Erklärung der Arbeitslosigkeit als diskretes Individualmerkmal:
0 = nicht arbeitslos, 1 = arbeitslos;
- die Analyse des Weiterbildungsverhaltens:
0 = keine Veranstaltungen besucht, 1 = Weiterbildungsveranstaltungen besucht;
- die Analyse sozialer Mobilität (Vergleich Berufsstatus des Vaters und des Sohnes):
0 = Berufsstatus vererbt, 1 = Berufsstatus verändert (Auf- oder Abstieg).

In der Regel setzt die Regressionsanalyse metrisch skalierte abhängige Merkmale voraus. Mit Hilfe von Dummy-Variablen gelingt jedoch auch eine Anwendung auf nominal skalierte Merkmale. Für eine bivariate Regression mit einer metrischen unabhängigen Variablen und einer 0/1-kodierten abhängigen Variablen erhalten wir

folgende Regressionsgleichung

Lineare Einfachregression (in Termen der i-ten Untersuchungseinheit):

$$B.1 \quad Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad i = 1, \dots, n$$

$$Y_1 = \beta_0 + \beta_1 X_1 + \varepsilon_1$$

$$Y_2 = \beta_0 + \beta_1 X_2 + \varepsilon_2$$

$$\vdots$$

$$\vdots$$

$$Y_n = \beta_0 + \beta_1 X_n + \varepsilon_n$$

Um spätere Verallgemeinerungen zu erleichtern und eine gedrängtere Darstellungsform zu erhalten, schreiben wir die Regression in Matrixform:

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$B.2 \quad \underset{n \times 1}{Y} = \underset{n \times 2}{X} \underset{2 \times 1}{\beta} + \underset{n \times 1}{\varepsilon}$$

Ausführlich geschrieben:

$$B.3 \quad \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 X_1 + \varepsilon_1 \\ \beta_0 + \beta_1 X_2 + \varepsilon_2 \\ \vdots \\ \beta_0 + \beta_1 X_n + \varepsilon_n \end{bmatrix}$$

$$d.h. \text{ wiederum: } Y_i = \beta_0 X_0 + \beta_1 X_i + \varepsilon_i \quad X_0 \equiv 1$$

Die Ausdrücke besagen im einzelnen:

\underline{Y} = Vektor der abhängigen (0/1)-Beobachtungswerte; \underline{X} = Vektor der unabhängigen

Variablenwerte: $\underline{\beta}$ = Vektor der Regressionskoeffizienten; $\underline{\epsilon}$ = Vektor der Vorhersagefehler.

Die Gleichung für die lineare Mehrfachregression mit allgemein k unabhängigen Variablen lautet dann:

$$\text{B. 4} \quad Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \epsilon_i$$

bzw. in Matrixform:

$$\text{B. 5} \quad \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \dots & x_{k1} \\ 1 & x_{12} & \dots & x_{k2} \\ \vdots & \vdots & & \vdots \\ 1 & x_{1n} & \dots & x_{kn} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

$$\underset{(n \times 1)}{\mathbf{y}} = \underset{(n \times k+1)}{\mathbf{X}} \underset{(k+1 \times 1)}{\boldsymbol{\beta}} + \underset{(n \times 1)}{\boldsymbol{\epsilon}}$$

wobei folgende Annahme für die Erwartungswerte der Fehlerterme gilt:

$$\text{B. 6} \quad E(\epsilon) = 0$$

$$E \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} = \begin{bmatrix} E(\epsilon_1) \\ E(\epsilon_2) \\ \vdots \\ E(\epsilon_n) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Bildet man nun z.B. in der Gleichung B.1 (bivariate Regression) auf der linken und der rechten Seite den Erwartungswert, so erhält man folgendes äquivalentes Regressionsmodell:

$$B.7 \quad \bar{E}(Y_i) = \beta_0 + \beta X_i, \quad i = 1, \dots, n$$

unter der Annahme von B.6.

Die Gleichung B.7 repräsentiert eine sogenannte lineare Wahrscheinlichkeitsfunktion. Da Y_i lediglich die Werte 0 und 1 annimmt, kann die Wahrscheinlichkeitsverteilung der Y_i wie folgt beschrieben werden:

Sei die Wahrscheinlichkeit, daß eine Untersuchungseinheit die Ausprägung "1" aufweist,

Daraus folgt: $\Pr(Y_i = 0) = 1 - \pi_i$ (als Komplementärwahrscheinlichkeit)
(PR = "probability").

Daraus folgt: $\Pr(Y_i = 1) = \pi_i$ (als Komplementärwahrscheinlichkeit)
(PR = "probability").

Der Ausdruck für den Erwartungswert von Y_i (Gleichung B.7) kann dann wie folgt erweitert werden:

$$\begin{aligned} E(Y_i) &= \pi_i(1) + (1 - \pi_i)(0) = \pi_i \\ \pi_i &= \alpha + \beta x_i \end{aligned}$$

d.h. der Erwartungswert von Y_i ist gleich der Wahrscheinlichkeit für $Y_i = 1$! Daraus folgt die Interpretation von Gleichung B.10 als lineares Wahrscheinlichkeitsmodell (11). Die Schätzwerte, die der Regressionsansatz liefert, sind als Wahrscheinlichkeit für das Ereignis " Y_i nimmt die Ausprägung 1 bei gegebenen Prädiktorwert x_i ", $i = 1, \dots, n$, zu interpretieren.

Es ist allerdings zu beachten, daß die rechte Seite der Gleichung B.7 größer als 1 oder kleiner als 0 sein kann; wenn für β die Schätzwerte $\hat{\beta}$ eingesetzt werden, während auf der linken Seite $E(Y)$ wegen $E(Y) \equiv P(Y \equiv 1)$ nicht außerhalb dieser Grenzen liegen kann. Tritt dieser Fall empirisch ein, so können die entsprechenden Ergebnisse nicht mehr sinnvoll interpretiert werden. Dies wird um so häufiger vorkommen, je näher die Wahrscheinlichkeit für das Auftreten der zu untersuchenden Merkmalsausprägung der abhängigen diskreten Variablen bei Null oder Eins liegt.

Eine Lösung des Problems besteht darin, die Werte für die Wahrscheinlichkeiten π in das interpretierbare 0/1-Intervall zu erzwingen, in dem die lineare Beziehung zwischen π und x innerhalb dieses Intervalls festgehalten wird:

$$\pi = \begin{cases} 1 & \beta_0 + \beta x > 1 \\ \beta_0 + \beta x & 0 \leq \beta_0 + \beta x \leq 1 \\ 0 & \beta_0 + \beta x < 0 \end{cases}$$

Dieser als "constrained linear probability model" bezeichnete Ansatz hat allerdings unattraktive Eigenschaften:

1. Unter bestimmten Bedingungen generiert dieses Modell instabile Schätzer.
2. Der abrupte Wechsel im Steigungskoeffizienten bei $TT = 0$ und $TT = 1$ führt zu Schwierigkeiten bei der Anpassung des Modells an die beobachteten Daten.
3. Die Wechsel selbst sind unter substantiellen Gesichtspunkten unwahrscheinlich.

Eine OLS-Schätzung ("ordinary least SQUARES") der unbekannten Modellparameter in dem linear-additiven Regressionsmodell wirft zwei weitere Probleme auf:

1. Da Y nur zwei Werte annehmen kann, können bei konstantem X die jeweiligen Fehler (ϵ_i) auch nur zwei Werte annehmen:

	ϵ_i	Wahrscheinlichkeit
$Y_i = 1$	$1 - \beta_0 - \beta X_i$	P_i
$Y_i = 0$	$-\beta_0 - \beta X_i$	$1 - P_i$

Die Fehler folgen daher auch nicht einer Normalverteilung, somit auch nicht die Schätzwerte für β_0 und β . Damit werden Signifikanztests sinnlos. Es läßt sich ferner zeigen, daß die Streuung der Störterme (ϵ) mit den Werten der unabhängigen Variablen variiert und die Homoskedastizitätsannahme somit nicht mehr haltbar ist:

Die Beziehung zwischen der Wahrscheinlichkeit P_i und X_i kann unter der Berücksichtigung unserer Annahmen, kann wie folgt bestimmt werden:

$$E(\epsilon_i) = (1 - \beta_0 - \beta X_i) P_i + (-\beta_0 - \beta X_i)(1 - P_i) = 0$$

Nach P_i aufgelöst erhalten wir:

$$P_i = \beta_0 + \beta X_i$$

$$1 - P_i = 1 - \beta_0 - \beta X_i$$

Die Varianz des Fehlerterms kann dann wie folgt berechnet werden:

$$\begin{aligned} E(\epsilon_i^2) &= (1 - \beta_0 - \beta X_i)^2 P_i + (-\beta_0 - \beta X_i)^2 (1 - P_i) \\ &= (1 - \beta_0 - \beta X_i)^2 (\beta_0 + \beta X_i) + (\beta_0 + \beta X_i)^2 (1 - \beta_0 - \beta X_i) \\ &= (1 - \beta_0 - \beta X_i)(\beta_0 + \beta X_i) = P_i(1 - P_i) \end{aligned}$$

$$\sigma_i^2 = E(\epsilon_i^2) = E(Y_i)[1 - E(Y_i)]$$

d.h. der Fehlerterm ist heteroskedastisch.

(Die OLS-Schätzer sind zwar noch erwartungstreu, aber nicht mehr effizient).

2. Das lineare Wahrscheinlichkeitsmodell läßt sich nicht auf den Fall polytomer abhängiger Merkmale verallgemeinern.

Die Probleme "Nichtnormalität und Heteroskedastizität der Störterme" sowie die Einschränkung der Anwendung auf dichotome abhängige Merkmale führten dazu, daß das lineare Wahrscheinlichkeitsmodell in zunehmenden Maße durch alternative Konzepte abgelöst wurde. In der Forschungspraxis verwendet man verschiedene Transformationen der abhängigen Variablen, die zu nicht-linearen Regressionsmodellen mit unterschiedlichen Verteilungsannahmen der Fehlerterme führen (12).

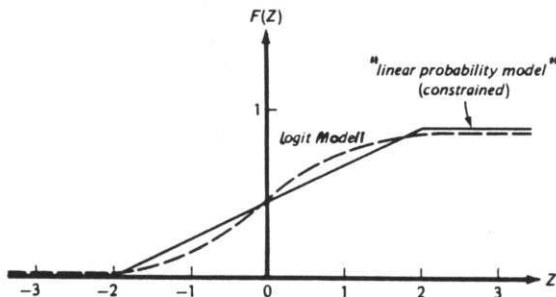
Verwendet man die logistische Verteilungsfunktion, so ergibt sich das Logit-Modell, verwendet man dagegen die Verteilungsfunktion der Normalverteilung, so spricht man von der Probit-Analyse. Als eine Alternative zur linearen Funktionsform wollen wir im folgenden lediglich auf das am weitaus häufigsten verwendete Logit-Modell eingehen.

B.1 DIE SPEZIFIKATION VON LOGIT-MODELLEN FÜR INDIVIDUALDATEN UND KONTINGENZTABELLEN

Zur Vereinfachung der Diskussion gehen wir im folgenden wiederum von einer dichotomen (0/1-kodierten) abhängigen Variablen aus, wobei wir zunächst Individualdaten annehmen und von einem bivariaten Erklärungsansatz ausgehen. Das Logit-Modell beruht auf der (kumulativen) logistischen Wahrscheinlichkeitsfunktion (13)

$$\text{B.1.1 } \pi_i(x_i) = P_i = F(Z_i) = \frac{1}{1 + e^{-Z_i}} = \frac{1}{1 + e^{-(\beta_0 + \beta x_i)}} \quad , \quad i = 1, \dots, n$$

wobei π_i die Wahrscheinlichkeit ist, daß die i -te Untersuchungseinheit bei gegebenem Prädiktorwert X_i in die Kategorie "1" der abhängigen Variablen fällt. Die Komplementärwahrscheinlichkeit $P(Y_i = 0)$ ist dann gleich $(1 - \pi_i)$. Die folgende Abbildung veranschaulicht die kumulative logistische Funktion:



$$\begin{aligned}
 1/P_i &= 1 + e^{-(\beta_0 + \beta x_i)} &= 1 + \exp(-\beta_0 - \beta x_i) \\
 (1/P_i) - 1 &= \frac{1 - P_i}{P_i} = e^{-(\beta_0 + \beta x_i)} &= \exp(-\beta_0 - \beta x_i) \\
 \frac{P_i}{1 - P_i} &= e^{\beta_0 + \beta x_i} &= \exp(\beta_0 + \beta x_i) \\
 1/P_i &= 1 + e^{-(\beta_0 + \beta x_i)} &= 1 + \exp(-\beta_0 - \beta x_i) \\
 (1/P_i) - 1 &= \frac{1 - P_i}{P_i} = e^{-(\beta_0 + \beta x_i)} &= \exp(-\beta_0 - \beta x_i) \\
 \frac{P_i}{1 - P_i} &= e^{\beta_0 + \beta x_i} &= \exp(\beta_0 + \beta x_i)
 \end{aligned}$$

Nach Logarithmieren der rechten und linken Gleichungshälfte folgt für die Schätzgleichung der Ausdruck

$$B.1.2 \quad \log \frac{P_i}{1 - P_i} = \beta_0 + \beta x_i, \quad i = 1, \dots, n \quad (\text{bivariate logistische Regression})$$

bzw.

$$B.1.3 \quad \log \frac{P_i}{1 - P_i} = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki}, \quad i = 1, \dots, n$$

für die multiple logistische Regression.

Der Ausdruck $\log(P_i/(1-P_i))$, der allgemein als LOGIT von π bezeichnet wird, beinhaltet das logarithmierte Verhältnis, daß Y_i eher den Wert "1" als den Wert "0" annimmt, bei gegebenen Prädiktorwerten $x_{1i} \dots x_{ki}$. Die abhängige Größe in der Regressionsgleichung ist in diesem Ansatz somit nicht mehr die ursprüngliche 0/1-kodierte Y-Variable, sondern die logarithmierten "odds".

Die Transformation des linearen Regressionsansatzes in eine logistische Funktionsform bewirkt, daß die transformierten Werte selbst zwischen $-\infty$ und $+\infty$ schwanken, während sich die geschätzten Wahrscheinlichkeiten in dem (0,1)-Wertebereich bewegen. Die Prognosegleichung für die Wahrscheinlichkeiten erhalten wir nach Umformung der Gleichung B.1.1:

$$P(Y_i=1) = \pi_i = \frac{1}{1 + e^{-Z_i}} = 1 - \frac{1}{1 + e^{Z_i}} = \frac{1 + e^{Z_i}}{1 + e^{Z_i}} - \frac{1}{1 + e^{Z_i}} = \frac{1 + e^{Z_i} - 1}{1 + e^{Z_i}},$$

daher:

$$B.1.4 \quad P(Y_i=1) = \pi_i = \frac{e^{Z_i}}{1 + e^{Z_i}} = \exp(Z_i) / (1 + \exp(Z_i)), \quad i = 1, \dots, n,$$

wobei (Z_i) die systematische Komponente $\beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki}$ darstellt.

Die Schätzung der unbekannten Regressionskoeffizienten in der systematischen Komponente erfolgt analog zum log-linearen Modell mit Hilfe der ML-Methode (Maximum-Likelihood-Schätzung).

Oft genug hat die diskrete abhängige Variable mehr als zwei Ausprägungen; man bezeichnet diese Variablen dann als polytom. Beschreiben wir die Anzahl der Ausprägungen eines diskreten Merkmals (Y) allgemein mit m, so können wir das Logit-Modell auf einfache Weise verallgemeinern, in dem wir zu der sogenannten multivariaten logistischen Verteilungsfunktion übergehen (/T4).

$$\text{B.1.5} \quad P(Y=j) = P_j = \frac{\exp(\underline{x}_j^T \underline{\beta}_j)}{\sum_{i=1}^m \exp(\underline{x}_i^T \underline{\beta}_i)} \quad j = 1, \dots, m$$

wobei \underline{x}^T der (k+1)-dimensionale (transponierte) Vektor der Einflußgrößen (Konstante und k unabhängige Variablen) ist:

$$\underline{x}^T = (1, x_1, \dots, x_k)$$

$$\text{und } \underline{\beta}_j = \begin{bmatrix} \beta_{0j} \\ \beta_{1j} \\ \vdots \\ \beta_{kj} \end{bmatrix} \quad (\text{Vektor der } k+1 \text{ Regressionskoeffizienten})$$

Da die Summe der Wahrscheinlichkeiten für jede Untersuchungseinheit über all m Kategorien der diskreten abhängigen Variablen gleich Eins ist, muß den β -Parametern eine sogenannte lineare Restriktion auferlegt werden. Mit anderen Worten: Das Modell B.1.5 ist nur dann eindeutig parametrisiert, wenn wir analog zur log-linearen Modellparametrisierung eine Normierung einführen, üblicherweise unterscheidet man zwischen einer

- symmetrischen Normierung (Restriktion: $\sum_{j=1}^m \beta_j = 0$) und die

- Referenzkategorie-Normierung (Restriktion: ein Parametervektor - entweder β_1 oder β_m - wird gleich Null gesetzt).

Um den Ausdruck B.1.5 besser verstehen zu können, sei noch einmal von einer dichotomen Zielvariable ausgegangen (d.h. m = 2). Unter der Annahme der symmetrischen Normierung folgt dann $\beta_1 = -\beta_2$. Ferner gilt: $p_1 = 1 - p_2$. Daraus folgt unmittelbar, daß

$$\log(P_1/P_2) = \log(P_1/(1-P_1)) = \mathbf{x}^T \beta_1 - \mathbf{x}^T \beta_2 = \mathbf{x}^T (\beta_1 - \beta_2) = \mathbf{x}^T (2\beta_1) \quad ,$$

d.h. $2\beta_1 = \beta$ wobei β der Parametervektor des dichotomen Logit-Modells aus der Gleichung B.1.3 ist! Wenn wir nun als Alternative die Restkategorie-Normierung mit Kategorie "2" als "baseline"-Ausprägung annehmen, (d.h. $\beta_m = \beta_2 = 0$), dann erhalten wir die Beziehung

$$\beta_1 = \beta$$

d.h. die Modellgleichung B.1.5 ist in diesem speziellen Fall identisch mit der Gleichung B.1.4! Allgemein gilt für $m > 2$ folgende Gegenüberstellung der Modell-Parametrisierungen:

SYMMETRISCHE NORMIERUNG	REFERENZKATEGORIE-NORMIERUNG
$\log(P_j/P_1) = \mathbf{x}^T (\beta_j - \beta_1)$	$\log(P_j/P_m) = \mathbf{x}^T \beta_j$
Die logarithmierten "odds" für jeweils ein Kategorienpaar sind eine lineare Funktion der Differenz zwischen zwei Parametervektoren.	Die Regressionskoeffizienten in dem Vektor β_j repräsentieren implizit jeweils die Effekte der unabhängigen Größen auf die "odds", daß eine Untersuchungseinheit in der Kategorie j relativ zur letzten Kategorie m prognostiziert wird.

Wir wollen im folgenden die Vorhersagegleichung für den allgemeinen Fall einer diskreten abhängigen Variablen mit m Ausprägungen unter Annahme der Referenzkategorie-Normierung ("baseline"-Ausprägung = m) aus den Schätzgleichungen des multinomialen Logit-Modells ableiten (vgl. Maddala 1983: S. 34f):

Sei Y eine abhängige diskrete Variable mit m Ausprägungen. Dann sind P_1, P_2, \dots, P_m die Wahrscheinlichkeiten für das Auftreten dieser Ausprägungen. Die grundlegende Idee ist nun wiederum, diese Wahrscheinlichkeiten in binärer Form (d.h. als "odds") auszudrücken. Seien

$$\begin{aligned}
 \text{B.1.6} \quad \frac{P_1}{P_1 + P_m} &= F(\beta'_1 x) \\
 \frac{P_2}{P_2 + P_m} &= F(\beta'_2 x) \\
 &\vdots \\
 \frac{P_{m-1}}{P_{m-1} + P_m} &= F(\beta'_{m-1} x)
 \end{aligned}$$

die (m-1) "odds". Sie implizieren die Wahrscheinlichkeitsverhältnisse

$$\text{B.1.7} \quad \frac{P_j}{P_m} = \frac{F(\beta'_j x)}{1 - F(\beta'_j x)} = G(\beta'_j x) \quad (j = 1, 2, \dots, m-1)$$

Da

$$\sum_{j=1}^{m-1} \frac{P_j}{P_m} = \frac{1 - P_m}{P_m} = \frac{1}{P_m} - 1$$

folgt die Gleichung

$$\text{B.1.8} \quad P_m = \left[1 + \sum_{j=1}^{m-1} G(\beta'_j x) \right]^{-1}$$

Daher folgt aus Gleichung B.1.7 die Vorhersagegleichung

$$\text{B.1.9} \quad P_j = \frac{G(\beta'_j x)}{1 + \sum_{j=1}^{m-1} G(\beta'_j x)}$$

Man geht davon aus, daß die Beobachtungswerte aus einer Multinomialverteilung mit den Wahrscheinlichkeiten (B.1.8) und (B.1.9) entstehen. Verwendet man eine logistische Verteilungsfunktion, so ist $G(\beta'_j x)$ aus Gleichung B.1.8 gerade $\exp(\beta'_j x)$ und die Gleichung B.1.8 sowie B.1.9 können wie folgt ausgedrückt werden:

$$\text{B.1.10} \quad P_j = e^{\beta'_j x} / D \quad (j = 1, 2, \dots, m-1) \quad ,$$

mit

$$P_m = 1/D$$

und

$$D = 1 + \sum_{k=1}^{m-1} e^{\beta'_k x}$$

Mit Gleichung B.1.10 haben wir die Grundgleichung des multinomialen Logit-Modells.

Damit können wir die Vorhersagegleichung des Logit-Modells für diskrete abhängige Merkmale mit allgemein m Ausprägungen zusammenfassend wie folgt ausdrücken:

$$P_{ij} = \frac{\exp(x'_i \beta_j)}{1 + \sum_{k=1}^{m-1} \exp(x'_i \beta_k)}$$

$$P_{im} = \frac{1}{1 + \sum_{k=1}^{m-1} \exp(x'_i \beta_k)}$$

mit $i = 1, \dots, n$ und $j = 1, \dots, m$

Mit der ausführlichen Darstellung des dichotomen und multivariaten Logit-Modells auf der Basis von Individualdaten wollten wir im Kontext der klassischen Regressionsanalyse deutlich machen, daß die Analyse diskreter abhängiger Merkmale ein direktes Analogon zu klassischen Verfahren der Abhängigkeitsanalyse darstellt. Nehmen wir z.B. in der Beobachtungsmatrix \underline{x} der unabhängigen Einflußgrößen neben metrischen Prädikatoren auch Dummy-Variablen auf, so hätten wir ein Pendant zur Kovarianzanalyse; hätte die Matrix \underline{x} nur Dummy-Variablen, so läge eine Varianzanalyse mit einer diskreten abhängigen Variablen vor. Ungewohnt ist zunächst die Logit-Transformation. Entscheidend für die Wahl eines bestimmten Erklärungsmodells ist vor allem, von welchen Verteilungsannahmen man über die abhängige Variable ausgeht und ob ein lineares Modell (= metrische Zielvariable) oder ein logistisches Modell (= diskrete abhängige Variable) bzgl. des empirischen Ausgangsmaterials angemessen ist. Die Interpretation der Werte für die Regressionskoeffizienten ist im Logit-Modell analog zu den klassischen Verfahren der Abhängigkeitsanalyse, wenngleich nicht so anschaulich aufgrund der logistischen Transformation.

Abschließend wollen wir zeigen, daß die für dichotome Zielvariablen dargestellte Modellstruktur (binomialverteilte Zielvariable, logistisches Regressionsmodell) erhalten bleibt, wenn wir das Aggregationsniveau der Daten ändern. Betrachten wir lediglich nominale/ordinale unabhängige Variablen, dann können wir analog zum log-linearen Modell die Ebene der Individualdaten ("case-level") verlassen und zu einem gruppenbezogenen Aggregationsniveau in Form einer mehrdimensionalen Kontingenztafel ("subclass level") übergehen. Voraussetzung wäre selbstverständlich auch in diesem speziellen Fall wiederum, daß eine hinreichend große Stichprobe vorliegt, so daß durch systematisches Kombinieren der Werte aller jeweils in eine Analyse einzubeziehenden Merkmale nicht zu häufig unbesetzte Zellen auftreten. Da im folgenden von einer asymmetrischen Fragestellung ausgegangen wird (d.h. wir führen eine Abhängigkeitsanalyse durch), ist die zu analysierende (mehrdimensionale) Kontingenztafel in spezifischer Weise zu formatieren. Das Beispiel auf der folgenden Seite mit drei dichotomen unabhängigen Merkmalen A, B und C sowie einer dichotomen abhängigen Variablen Y illustriert den Aufbau einer vierdimensionalen Tabelle für eine Abhängigkeitsanalyse.

Die Zusammenfassung der Individualdaten zu einer aggregierten Datenbasis ergibt bzgl. der unabhängigen Merkmale A, B und C maximal $2 \times 2 \times 2 = 8$ Subpopulationen, da auch nur acht verschiedene Merkmalskombinationen auftreten können. Die Häufigkeit, mit der die einzelnen Merkmalskombinationen der unabhängigen Merkmale auftreten können, bezeichnen wir allgemein mit (n_i) . Neben der Gesamthäufigkeit pro Subpopulation stellen wir noch die konditionale (oder bedingte) Verteilung des abhängigen Merkmals Y fest. Mit anderen Worten: Für jede Subpopulation i ($i = 1, \dots, 8$) stellen

SUBPOPULATION (i)	A	B	C	Y = 0 n(0)	Y = 1 n(r)	P(Y=1) (P _i)	GESAMTHÄUFIGKEIT n _i
1	1	1	1	n ₀₁	n ₁₁	P ₁	n ₁
2	1	1	2	n ₀₂	n ₁₂	P ₂	n ₂
3	1	2	1	n ₀₃	n ₁₃	P ₃	n ₃
4	1	2	2	n ₀₄	n ₁₄	P ₄	n ₄
5	2	1	1	n ₀₅	n ₁₅	P ₅	n ₅
6	2	1	2	n ₀₆	n ₁₆	P ₆	n ₆
7	2	2	1	n ₀₇	n ₁₇	P ₇	n ₇
8	2	2	2	n ₀₈	n ₁₈	P ₈	n ₈
$r_i = n_{1i} \quad ; \quad P_i = r_i/n_i \quad ; \quad 1 - P_i = 1 - r_i/n_i$							

wir fest, wieviele der jeweils n_i Untersuchungseinheiten die Ausprägung $Y = 1$ bzw. $Y = 0$ aufweisen. Diese Vorgehensweise führt zu den bedingten Verteilungen n_{0i} und n_{1i} (im folgenden mit r_i abgekürzt):

Das logistische Modell ist auch bei aggregierten Daten ein sinnvoller Erklärungsansatz, wenn die Subpopulation (i) selbst die Untersuchungseinheiten darstellen und eine geeignete "Metrisierung" der abhängigen Variablen definiert wird. Im Gegensatz zum GSK-Modell werden im logistischen Regressionsansatz im Falle aggregierter Daten aus den konditionalen Häufigkeiten des abhängigen Merkmals die bekannten logarithmierten "odds" für jede Subpopulation (i) gebildet:

$$\log \left(\frac{r_i/n_i}{1 - r_i/n_i} \right) = \log \left(\frac{r_i}{n_i - r_i} \right)$$

Zur Illustration der logistischen Regression auf der Basis von aggregierten Daten gehen wir von einer dreidimensionalen Kontingenztafel aus (s. folgende Seite). Es sei auch für dieses Beispiel explizit darauf hingewiesen, daß wir die ursprüngliche Variablenausprägungen auf die angegebenen Kategorien reduziert haben, um das Datenmaterial übersichtlich zu halten. Unter inhaltlichen Gesichtspunkten ist diese Vorgehensweise sicherlich unbefriedigend, aber zur Illustration der Methode wird auch

STADT	BILDUNG	MOB		TOTAL
		IMMOBILE	MOBILE	
MINDEN	VOLKSSCH	1432	554	1986
	MITTTERT	686	2403	3089
	SEKPRIMA	183	156	339
	TOTAL	2301	3113	5414
DUISBURG	VOLKSSCH	6278	2294	8572
	MITTTERT	362	596	958
	SEKPRIMA	138	188	326
	TOTAL	6778	3078	9856

VARIABLEN DES ANWENDUNGSBEISPIELS:

MOB (=Y): 0 = Berufsstatus der Söhne hat sich gegenüber dem der Väter nicht verändert (immobile)
 1 = Berufsstatus hat sich verändert (ohne Unterscheidung nach Abstiegen bzw. Aufstiegen)

BILDUNG: 1 = Volksschule
 (SOHN) 2 = Mittelschule bis Tertia
 3 = Sekunda bis Prima

STADT: 1 = Minden
 2 = Duisburg

in diesem Fall ein Kompromiß im Hinblick auf Anschaulichkeit gemacht. Da lediglich ein nominales und ein ordinales Merkmal als unabhängige Größen zur Erklärung des Ereignisses "mobil" ($Y = 1$) im Verhältnis zum Ereignis "immobil" ($Y = 0$) herangezogen wird, ist zunächst (analog zum log-linearen Modell) eine geeignete Design-Matrix zu konstruieren. Eine unabhängige Variable mit allgemein k Ausprägungen kann im Analysedesign durch maximal $(k-1)$ Dummy-Vektoren erfaßt werden, da die Design-Matrix \underline{x} vollen Spaltenrang aufweisen muß (d.h. es dürfen keine linearen Abhängigkeiten in \underline{x} auftreten). Die Lösung des Problems besteht wiederum in der Formulierung der bereits im ersten Abschnitt diskutierten Reparametrisierungsbedingungen:

- Effekt-Kodierung oder
- Dummy-Kodierung.

Beide Techniken unterscheiden sich lediglich in der Behandlung der 'Basiskategorie', die in der Design-Matrix nicht durch einen eigenen Vektor abgebildet wird: sie wird im ersten Fall durchgängig mit "-1" und im zweiten Fall durchgängig mit "0" kodiert; am Beispiel einer qualitativen Variablen A mit drei Ausprägungen und Kategorie 1 als Basiskategorie:

'Gruppen'	Erfassung der 'Gruppenzugehörigkeit' über			
	EFFEKT-KODIERUNG		DUMMY-KODIERUNG	
	Vektoren		Vektoren	
	A_2	A_3	A_2	A_3
A(1)	-1	-1	0	0
A(2)	1	0	1	0
A(3)	0	1	0	1

Auf der folgenden Seite haben wir die Design-Matrizen für unser exemplarisches Beispiel mit den unterschiedlichen Kodierungsformen vergleichend gegenübergestellt, wobei wir lediglich jeweils die Haupteffekte der unabhängigen Variablen berücksichtigt haben. Die im Rahmen der logistischen Regression berechneten Koeffizienten haben wiederum eine unterschiedliche Bedeutung (s. Abschnitt A): Im Falle der Dummy-Kodierung werden 'cornered effects' und im Falle der Effekt-Kodierung 'centered effects' berechnet. Das bedeutet, daß die erzielten Regressionskoeffizienten im ersten Fall jeweils relativ zur Basiskategorie und im zweiten Fall als Abweichung vom 'Grand Mean' der abhängigen Variablen (= logarithmierte "odds"), dem Grundeffekt, berechnet werden. Jede Spalte in der Design-Matrix repräsentiert somit eine unabhängige Variable, die jeweils den Effekt der korrespondierenden Ausprägung eines Merkmals quantifiziert.

Bevor wir die Regressionskoeffizienten eines bestimmten Modells diskutieren, ist zunächst einmal ein sparsameres, aber dem empirischen Befund noch hinreichend angepaßtes Modell zu suchen. Die Beobachtungswerte der 'metrisierten' abhängigen Variablen (d.h. die logarithmierten "odds") in den (i) Subpopulationen sind durch wenige unabhängige Variablen (d.h. Spalten einer Design-Matrix) zu erklären. Das 'optimale' Modell mit hoher Erklärungskraft und wenigen Regressionskoeffizienten wird innerhalb einer Hierarchie unterschiedlicher Modellhypothesen liegen, wobei analog zum log-linearen Modellansatz das Basis-Modell und das saturierte Modell die extremen Endpunkte der Hierarchie bilden:

- Null-Modell (M_0)
- .
- .
- .
- Aktuelles Modell (M_c) mit p Regressionskoeffizienten
- .
- .
- .
- Saturiertes Modell (M_s) mit n Regressionkoeffizienten

Für die Entscheidung zugunsten eines bestimmten Modells innerhalb dieser Hierarchie verwendet man die Theorie des Likelihood-Verhältnis-Tests zur Beurteilung der Güte der Modellanpassung. Dabei wird der Wert der Likelihood-Funktion eines eingeschränkten Modells (L_0) mit dem Wert der Likelihood-Funktion eines weitergehenden Modells (L_1) verglichen, wobei die Parameter des eingeschränkten Modells in dem weitergehenden Modell enthalten sein müssen ("nested models"). Die Likelihood-Ratio-Teststatistik lautet dann

$$B.1.11 \quad \chi^2_L = G^2 = -2 \log(L_0/L_1) = 2 (\log L_1 - \log L_0)$$

DESIGN-MATRIZEN UNTER ANNAHME ALTERNATIVER REPARAMETRISIERUNGS-BEDINGUNGEN

SUBPOPULATION	(S) (B)		DESIGN-MATRIX MIT ZENTRIERTEN EFFEKTEN			
			STADT	BILDUNG(2)	BILDUNG(3)	KONSTANTE
1	1	1	-1	-1	-1	1
2	1	2	-1	1	0	1
3	1	3	-1	0	1	1
4	2	1	1	-1	-1	1
5	2	2	1	1	0	1
6	2	3	1	0	1	1

SUBPOPULATION	(S) (B)		DESIGN-MATRIX MIT "CORNERED EFFECTS"			
			STADT	BILDUNG(2)	BILDUNG(3)	KONSTANTE
1	1	1	0	0	0	1
2	1	2	0	1	0	1
3	1	3	0	0	1	1
4	2	1	1	0	0	1
5	2	2	1	1	0	1
6	2	3	1	0	1	1

Die ersten Ausprägungen der Variablen STADT und BILDUNG dienen jeweils jeweils als Basiskategorie.

DEFINITION DER STELLVERTRETERVARIABLEN

1. Zentrierte Effekte:

$$B(2) = \begin{cases} -1 & \text{für BILDUNG} = 1. \text{ Kategorie (Basiskat.)} \\ +1 & \text{" " " = 2. " "} \\ 0 & \text{" sonstige Kategorie} \end{cases}$$

$$B(3) = \begin{cases} -1 & \text{für BILDUNG} = 1. \text{ Kategorie (Basiskat.)} \\ +1 & \text{" " " = 3. " "} \\ 0 & \text{" sonstige Kategorie} \end{cases}$$

$$S(2) = \begin{cases} -1 & \text{für STADT} = 1. \text{ Kategorie (Basiskat.)} \\ +1 & \text{" " " = 2. " "} \end{cases}$$

2. Auf eine Kategorie bezogene Effekte:

$$B(2) = \begin{cases} +1 & \text{für BILDUNG} = 2. \text{ Kategorie} \\ 0 & \text{sonst} \end{cases} \quad B(3) = \begin{cases} +1 & \text{für BILDUNG} \\ & = 3. \text{ Kategorie} \\ 0 & \text{sonst} \end{cases}$$

$$S(2) = \begin{cases} +1 & \text{für STADT} = 2. \text{ Kategorie} \\ 0 & \text{sonst} \end{cases}$$

Führen wir einen Anpassungstest durch, dann vergleichen wir das jeweils aktuelle Modell "constrained model") mit dem saturierten Modell ("füll model"). Nach Gleichung B.1.11 ergibt sich dann folgender Ausdruck für das Likelihood-Verhältnis:

$$B.1.12 \quad G_0^2(q) = -2 \log(L_c/L_f) = 2 (\log L_f - \log L_c)$$

die mit d.f. = $q = n - p$ Freiheitsgraden asymptotisch χ^2 -verteilt ist. Der Vergleich mit dem kritischen Wert der χ^2 -Verteilung darf in diesem speziellen Fall keine signifikanten Ergebnisse liefern!

Vergleichen wir dagegen den Likelihood-Wert des Null-Modells M^{\wedge} (= eingeschränktes Modell) mit dem Likelihood-Wert eines aktuellen Modells (= weitergehendes Modell), so muß die Teststatistik nach B.1.11 ein signifikantes Ergebnis liefern, will man das Modell beibehalten! Hier handelt es sich also wieder um einen üblichen Signifikanztest (analog zum F-Test in der klassischen Regressionsanalyse). Partielle Test, mit deren Hilfe wir einzelne Effekte oder Effekteilmengen auf Signifikanz überprüfen können, resultieren aus dem Vergleich der Likelihood-Werte aus zwei geschachtelten (hierarchischen) aktuellen Modellen, die sich gerade um die (oder den) zu testenden Parameter unterscheiden:

$$B.1.13 \quad G_{Part}^2(1) = -2 \log(L_{M_{c1}}/L_{M_{c2}}) = 2 (\log L_{M_{c2}} - \log L_{M_{c1}})$$

Die Teststatistik ist ebenfalls χ^2 -verteilt mit d.f. = $1 = p_1 - p_2$ Freiheitsgraden (p_1 = Anzahl der Parameter im weitergehenden Modell; p_2 = Anzahl der Parameter im eingeschränkten Modell).

Die Logik dieses Tests verlangt, daß sowohl das Modell M_{c1} und das Modell M_{c2} den Daten hinreichend angepaßt ist (die Anpassungstests dürfen somit nicht signifikant sein). Ferner ist zu beachten, daß es sich hier um ein Pendant zum partiellen F-Test in der Regressionsanalyse handelt, d.h. wir sind an signifikanten Ergebnissen interessiert. Mit Hilfe der partiellen Teststatistik kann man z.B. fragen, ob die Vernachlässigung mehrerer Parameter die Modellanpassung signifikant verschlechtert. Wenn der partielle Wert des reduzierten Modells gegenüber dem Ausgangsmodell einen bestimmten kritischen Wert überschreitet, d.h. signifikant ist, dann kann man davon ausgehen, daß der oder die jeweiligen Parameter im Modell verbleiben müssen. Die Diskussion zeigt, daß die Test-Statistiken des logistischen Regressionsmodells mit jenen des log-linearen Modellansatzes identisch sind, denn sie sind für beide Modellvarianten das Ergebnis der Maximum-Likelihood-Schätzung;

Analog zum log-linearen Modell können daher auch multiple und partielle Bestimmtheitsmaße auf der Basis von Devianzen ($G_0^2(q)$ -Werten) berechnet werden:

$$B.1.14 \quad R^2_{CHI^2_L}(x_1 \dots x_p) = \frac{(CHI^2_L - M_0 - CHI^2_L - M_c)}{CHI^2_L - M_0} = 1 - \frac{(CHI^2_L - M_c)}{CHI^2_L - M_0}.$$

Diese Maßzahl gibt (multipliziert mit 100) den Prozentanteil reduzierter Devianz eines aktuellen Modells relativ zum Basismodell an.

$$B.1.15 \quad PRD_{CHI^2_L}(x_i) = \frac{(CHI^2_L - M_{c1} - CHI^2_L - M_{c2})}{CHI^2_L - M_{c1}}.$$

PRD gibt an, um welchen Prozentsatz die Devianz in dem Modell M_{c1} durch Einführung der zusätzlichen Variablen (x_i) in dem Modell M_{c2} verringert wird. Die Modelle M_{c1} und M_{c2} unterscheiden sich also gerade in der Variablen x_i .

$$B.1.16 \quad sr^2_{CHI^2_L}(x_i) = \frac{(CHI^2_L - M_{c1} - CHI^2_L - M_{c2})}{CHI^2_L - M_0}$$

Die Maßzahl $sr^2(x_i)$ entspricht gerade dem Anteil erklärter Devianz, der auf die Variable (x_i) zurückzuführen ist. Die folgende Tabelle enthält die diskutierten Statistiken und Maßzahlen für eine Sequenz von logistischen Modellen zu dem Beispiel auf S. 55

HIERARCHISCHE LOGISTISCHE MODELLE AUF DER BASIS VON AGGREGIERTEN DATEN								
MODELL	EFFEKT DER VARIABLEN	DEVIANZ (G^2_o)	d.f.	p	R^2	PRD (x_i)	$sr^2(x_i)$	log (L)
BM	-	2856	5	0.0	0.0	0.0	0.0	-10309
B + S	B,S	73	2	0.0	0.97	0.97	0.97	- 8918
B	S	98	3	0.0	0.96	0.25	0.01	- 8930
S	B	1860	4	0.0	0.35	0.96	0.62	- 9811
B+S+B*S = sat. Mod.	B,S,B*S	0	0	1.0	1.00	-	-	- 8882
Abkürzungen: BM = Basismodell; B = Bildung; S = Stadt; B*S = Interaktion BILDUNGxSTADT								

Nach dem Signifikanzkriterium müßten wir das saturierte Modell akzeptieren, das die beobachteten Häufigkeiten perfekt reproduziert. Diese Entscheidung überrascht uns nicht, denn vor dem Hintergrund der großen Fallzahl ($N = 15270!$) war dieses Ergebnis

zu erwarten. Der χ^2 -Test reagiert - wir hatten bereits darauf hingewiesen - äußerst sensibel auf große Stichprobenumfänge. Wir stützen daher unsere Entscheidung zugunsten eines sparsameren Modells ausschließlich auf die deskriptiven Maßzahlen der Modellanpassung:

- Das einfachste Modell, das den empirischen Befund noch angemessen beschreibt, ist das Modell mit den Effekten der Variablen BILDUNG.
- Wird neben der Variablen STADT zusätzlich die Variable BILDUNG in dem Erklärungsansatz berücksichtigt, so verringert sich die Devianz um 96% (dagegen lediglich um 25%, wenn man neben BILDUNG zusätzlich STADT berücksichtigt).
- Betrachtet man das Haupteffektmodell (B + S), so ergibt sich für die Variable BILDUNG ein relativer Erklärungsanteil von 62% und für STADT ein relativer Anteil von 1%.

Die Ergebnisse stützen somit zusammen betrachtet die Entscheidung zugunsten des Modells mit der unabhängigen Variablen BILDUNG. Es gibt somit keinen stadtsspezifischen Einfluß auf den Mobilitätsgrad.

Der hohe Wert für das Bestimmtheitsmaß (Modell mit B: $R^2 = 0.96$) darf nicht dazu verführen, in der Variablen BILDUNG einen besonders erklärungskräftigen Prädiktor zu sehen. Diese Aussage wäre sicherlich eine erhebliche Überinterpretation, denn die multiplen Bestimmtheitsmaße für Modellsequenzen werden nach dem oben dargestellten Gleichungen ausschließlich über χ^2_L -Differenzen relativ zum Basismodell berechnet. Danach "erklärt" ein saturiertes Modell ($\chi^2_L = 0$) den empirischen Befund perfekt. Die Analogie zum $R^2_{y.x_1 \dots x_p}$ der klassischen multiplen Regression auf der Basis von Individualdaten trägt somit nicht sehr weit. Ein wesentlicher Grund für die Größenordnungen des Bestimmtheitsmaßes ist in der Aggregation der Individualdaten zu Kontingenztabellen zu sehen: Die Größe eines $R^2_{\chi^2_L}$ -Wertes hängt - bei gegebener Dimension der jeweils zugrundeliegenden Kontingenztafel - allein von der Anzahl der spezifizierten Effekte ab. Eine logistische Regressionsanalyse auf der Basis von Individualdaten liefert dagegen R^2 -Werte, deren Größenordnungen direkt mit den R^2 -Werten einer multiplen Regression mit metrischer Zielvariablen vergleichbar sind.

Nachdem ein sparsames Modell qualifiziert ist, sind die Schätzwerte für die einzelnen Parameter zu interpretieren. Wir gehen im folgenden von dem Haupteffektmodell aus, obwohl die Variable STADT keinen nennenswerten Einfluß auf den Mobilitätsgrad hat (einzig darstellerische Gründe sprechen in diesem Fall dafür!). In der Tabelle auf der folgenden Seite sind die Parameterschätzwerte, ihre Standardabweichungen und die Bezeichnungen der Effekte für die zwei diskutierten Reparametrisierungsbedingungen gegenübergestellt. Die aus der Varianz-Kovarianzmatrix der berechneten Schätzer $\hat{\beta}_j$ resultierenden Standardfehler (= Wurzel aus den Varianzen der Schätzer $\hat{\beta}_j$) sind eine

Maßzahl für die Genauigkeit der Schätzung. Sie können für Hypothesentests einzelner Parameterwerte herangezogen werden. Will man z.B. überprüfen, ob eine unabhängige Einflußgröße x_j der Design-Matrix \underline{x} einen von Null verschiedenen Effekt auf die Zielgröße hat, so ist mit einer (zuvor festgelegten) Sicherheit α (= Sicherheitswahrscheinlichkeit) auszuschließen, daß der Wert unter Annahme der Nullhypothese $\beta_j = 0$

in dem sogenannten zweiseitigen Konfidenzintervall $(1 - \alpha)$ enthalten ist. Bei der praktischen Durchführung gehen wir von folgendem Ausdruck aus:

$$\text{B.1.17} \quad t = \hat{\beta}_j / \sqrt{V(\beta_j)}, \quad \text{wobei } \sqrt{V(\beta_j)} = \text{Standardfehler von } \beta_j.$$

Ist dieser Quotient größer als der kritische Wert der Standardnormalverteilung (Z),

$$t > Z_{1 - \alpha/2},$$

dann wird die Nullhypothese ($\beta_j = 0$) verworfen. Als Faustregel gilt:

Ein Regressionskoeffizient ist auf dem Alpha-Niveau von 5% signifikant, dessen Wert mindestens doppelt so groß ist wie sein Standardfehler.

In der Tabelle auf S.62 haben wir neben den Parameterschätzwerten jeweils den Standardfehler aufgeführt. Betrachten wir zunächst einmal die Regressionskoeffizienten der Effekt-Kodierung:

- Die Terme S(2) und B(2) sind auf dem 5%-Niveau signifikant. Unter substanziellen Gesichtspunkten haben wir allerdings festgestellt, daß die Variable STADT keinen nennenswerten Einfluß auf das (logarithmierte) Verhältnis Mobile/Immobilie hat.
- Das als "Konstante" gekennzeichnete absolute Glied der Regressionsgleichung besagt, daß der Durchschnitt der logarithmierten "odds" gerade 0.04 beträgt.
- An der Form der Effekt-Kodierung (s.S. 57) erkennt man, daß der β -Wert von S(2) als Abweichung vom Gesamtdurchschnitt zu interpretieren ist, d.h. der Durchschnittswert für die Ausprägung 'Duisburg' (mit +1 in der Design-Matrix kodiert) sinkt von 0.04 auf -0.07 ($0.04 - 0.11 = -0.07$). Mit anderen Worten: Der durchschnittliche Anteil der Mobilen verringert sich gegenüber der Stadt Minden (Durchschnittswert der "log-odds" in Minden: $0.04 + 0.11 = 0.15$). Die Spanne zwischen den beiden Kategorien einer unabhängigen Variablen ist also immer gleich dem doppelten Wert des Regressionskoeffizienten.
- Sehr viel bedeutsamer ist der Effekt der Schulbildung S(2) auf das Verhältnis Mobile/Immobilie: Jene Söhne, die in der Kategorie "Mittelschule-Tertia" beobachtet wurden, weisen einen deutlich höheren Mobilitätsgrad auf als etwa die Gruppe der Volksschüler oder die Söhne in der Kategorie B(3) "Sekunda-Prima".

Für B = 1 erhalten wir $0.04 - 0.97 = -0.93$,

für B = 2 erhalten wir $0.04 + 0.95 = 0.99$ und

für B = 3 erhalten wir $0.04 + 0.02 = 0.06$.

SCHÄTZWERTE FÜR DIE PARAMETER DES HAUPTEFFEKTMODELLS (B+S) AUF DER BASIS AGGREGIERTER DATEN

EFFEKT-KODIERUNG				DUMMY-KODIERUNG			
Parameter	Schätzwert	Standard- fehler	Wert/S.E.	Parameter	Schätzwert	Standard- fehler	Wert/S.E.
S(2)	-0,11	0,022	-5,0	S(2)	-0,22	0,04	-5,0
B(2)	0,95	0,038	25,2	B(2)	1,92	0,05	39,9
B(3)	0,02	0,05	0,4	B(3)	0,99	0,08	12,2
Konstante	0,04	0,03	1,5	Konstante	-0,82	0,04	-19,9

ZUSAMMENFASSENDE BESCHREIBUNG DER KONTINGENZTABELLENZELLEN							
Subpopu- lation	STADT	BILDUNG	Mobile	Immobilie	P_i (Mobile)	\hat{P}_i (Mobile)	Geschätzte log odds
1	1	1	554	1432	0,279	0,306	-0,819
2	1	2	2403	686	0,778	0,751	1,106
3	1	3	156	183	0,460	0,544	0,177
4	2	1	2294	6278	0,267	0,261	-1,039
5	2	2	596	362	0,622	0,708	0,88
6	2	3	188	138	0,577	0,489	-0,043

Der Effekt der Kategorie B(1) (-0.97) folgt aus der Reparametrisierungsbedingung "Summe aller Werte über ein Merkmal gleich Null", d.h.

$$\begin{array}{ccccccc} \text{Effekt B(1)} & + & \text{Effekt B(2)} & + & \text{Effekt B(3)} & = & \text{Null} \\ -0.97 & + & 0.95 & + & 0.02 & = & 0 \end{array}$$

Da wir in diesem Modell keine Interaktion (oder Wechselwirkung) zwischen STADT und BILDUNG annehmen und so auch nicht in der Design-Matrix spezifiziert haben (als Produkt der Spalten S(2), B(2) und B(3): $S(2) \cdot B(2)$ und $S(2) \cdot B(3)$), unterscheiden sich die Effekte der Variablen BILDUNG in Minden und Duisburg nicht voneinander. Wir gehen somit nicht von einem stadtspezifisch variierenden Effekt der BILDUNG auf den Mobilitätsgrad aus.

Die Interpretation der einzelnen Effekte als Abweichungen vom Gesamtdurchschnitt kann leicht anhand der logarithmierten "odds" als "metrisierte" Zielvariable in der unteren Tabelle auf Seite 62 nachvollzogen werden. Für die einzelnen Ausprägungskombinationen der unabhängigen Merkmale haben wir schließlich noch die beobachteten Anteilswerte (= Wahrscheinlichkeiten) und die modellimplizierten (geschätzten) Anteilswerte für die Ausprägung "mobil" der abhängigen Variablen aufgeführt (15) Die Differenzen zwischen diesen beiden Anteilswerten sind durchgängig nicht sehr groß. Dieses Ergebnis bestätigt wiederum die gute Anpassung des Haupteffektmodells an den empirischen Befund.

Wenden wir uns anschließend noch kurz der Effekt-Interpretation unter Verwendung der Dummy-Kodierung zu. Als Reparametrisierungsbedingung haben wir die Effekte der jeweils ersten Ausprägung einer Variablen gleich Null gesetzt. Die erste Kategorie dient jeweils als Bezugs- oder Basiskategorie. Daraus läßt sich folgende Interpretation der Effekte S(2), B(2) und B(3) ableiten:

- Die Regressionskonstante gibt - nach Transformation (vgl. Gl. B.1.4) - die Wahrscheinlichkeit an, daß Söhne mit Volksschulabschluß (= erste Kategorie der Variablen BILDUNG) aus Minden (= erste Kategorie der Variablen STADT) mobil waren:

$$\begin{aligned} p_{11}^{SB} &= \exp(\text{Konstante}) / (1 + \exp(\text{Konstante})) \\ &= \exp(-0.82) / (1 + \exp(-0.82)) = 0.306 \end{aligned}$$

- Im Vergleich zu B(1) - nur Volksschulabschluß - wird die Wahrscheinlichkeit, mobil zu sein, erhöht, wenn Söhne eine mittlere Bildung (Mittelschule-Tertia; B(2) = 1.92) oder eine höhere Bildung (Sekunda - Prima; B(3) = 0.99) besitzen. Allerdings ist der Effekt der Kategorie B(3) längst nicht so bedeutsam wie der Effekt der zweiten Ausprägung (im Vergleich jeweils zur ersten Ausprägung).
- Die Variable STADT beeinflusst den Mobilitätsgrad wie folgt: Für Söhne aus Duisburg vermindert sich die Wahrscheinlichkeit, mobil zu sein, geringfügig.

- Die Werte der Effekte sind unter Annahme der Dummy-Kodierung jeweils relativ zur Basiskategorie zu interpretieren. Die folgenden Beispiele veranschaulichen die Interpretationsrichtung der "cornered effects":

$$(a) \quad \log \text{ odd} \left(\overset{SB}{21} \right) - \log \text{ odd} \left(\overset{SB}{11} \right) = \text{Effekt } S(2) \\ (-1.039) - (-0.819) = -0.22$$

$$\log \text{ odd} \left(\overset{SB}{22} \right) - \log \text{ odd} \left(\overset{SB}{12} \right) = \text{Effekt } S(2) \\ (0.88) - (1.106) = -0.22$$

$$(b) \quad \log \text{ odd} \left(\overset{SB}{13} \right) - \log \text{ odd} \left(\overset{SB}{11} \right) = \text{Effekt } B(3) \\ (0.177) - (-0.819) = 0.99$$

$$\log \text{ odd} \left(\overset{SB}{23} \right) - \log \text{ odd} \left(\overset{SB}{21} \right) = \text{Effekt } B(3) \\ (-0.043) - (-1.309) = 0.99$$

Die Beispiele zeigen, daß die Effekte der Variablen **BILDUNG** nicht stadtsspezifisch variieren. Umgekehrt gilt: Der Effekt der Variablen **STADT** variiert nicht mit den Ausprägungen der Variablen **BILDUNG**. Das liegt wiederum daran, daß wir in der Design-Matrix keine Interaktionen in Form von Produktvariablen spezifiziert haben (wir gehen von einem Haupteffekt-Modell aus).

- Die folgenden Beispiele illustrieren die Anwendung der Prognosegleichung:
 - (a) Wahrscheinlichkeit, mobil zu sein, für Söhne aus Duisburg mit einem Volks schulabschluß:

$$P_{21}^{SB} = \exp(-0.82 + (-0.22)) / (1 + \exp(-0.82 + (-0.22))) = 0.261;$$

- (b) Wahrscheinlichkeit, mobil zu sein, für Söhne aus Duisburg mit mittlerer Bildung (Mittelschulabschluß bis Tertia):

$$P_{22}^{SB} = \exp(-0.82 + (-0.22) + 1.92) / (1 + \exp(-0.82 + (-0.22) + 1.92)) = 0.708$$

Insgesamt gesehen ist die Interpretation der Effekte unter einer bestimmten Reparametrisierungsbedingung völlig analog zur klassischen Varianzanalyse bzw. Regressionsanalyse mit Stellvertretervariablen zu bewerkstelligen. Ungewohnt mag zunächst die Logit-Transformation sein: Die Verwendung von Logits als "metrisierter" Zielvariablen ist nicht sehr anschaulich. Während sich die Effektinterpretation im Rahmen der logistische Regression auf logarithmierte "odds" bezieht, erfolgt die Effektinterpretation im Rahmen der Varianzanalyse anhand von Mittelwertvergleichen. Dennoch sehen wir in dem logistischen Regressionsansatz - sowohl auf der Basis von Indivi-

dualdaten als auch auf der Basis von aggregierten Daten (Kontingenztabellen) - ein geeignetes Verfahren zur Analyse diskreter abhängiger Merkmale.

Mit der hier verkürzt illustrierten Form der logistischen Regression haben wir zu dem nur einen Teil der Analysemöglichkeiten ausgeschöpft. So ist es z.B. im Rahmen einer Kontingenztabellenanalyse möglich, neben den Effekten nominaler/ordinaler Verfahren auch metrische unabhängige Merkmale in der Design-Matrix aufzunehmen. Gegeben sei eine dichotome abhängige Variable Y; neben den nominalen Prädiktoren A(1,2) und B(1,2) sollen in einer Regressionsanalyse auf der Basis von aggregierten Daten auch metrische Prädiktoren x_1 und x_2 herangezogen werden. Wir spezifizieren dann folgende Design-Matrix.

Subpopulation	A	B	A(2)	B(2)	x_1	x_2
1	1	1	0	0	\bar{x}_{11}	\bar{x}_{12}
2	1	2	0	1	\bar{x}_{21}	\bar{x}_{22}
3	2	1	1	0	\bar{x}_{31}	\bar{x}_{32}
4	2	2	1	1	\bar{x}_{41}	\bar{x}_{42}

Für jede unabhängige metrische Variable wird ihr Mittelwert pro Subpopulation (gebildet über die Ausprägungskombinationen der nominalen Merkmale) berücksichtigt. Ein anschauliches Beispiel für diesen Typ findet man in Arminger (1983a). Es sei jedoch explizit darauf hingewiesen, daß man bzgl. der metrischen Merkmale aufgrund der Datenaggregation (Berechnung von subpopulationsspezifischen Mittelwerten) gewisse Informationsverluste in Kauf nehmen muß. Betrachtet man in einer Untersuchung eine hohe Anzahl metrischer Prädiktoren und z.B. nur wenige nominale unabhängige Merkmale, dann ist eine logistische Regression auf der Basis von Individualdaten sicherlich angemessener.

B.2 LOG-LINEARE MODELLBILDUNG FÜR FRAGESTELLUNGEN MIT EXPLIZITER ABHÄNGIGER VARIABLEN

Dem in den ersten Abschnitt (A) dargestellten log-linearen Modell liegt eine symmetrische Betrachtungsweise zugrunde; jedes Merkmal wird in gleicher Weise behandelt, da lediglich die Zusammenhangsstruktur zwischen ausgewählten Variablen in einer mehrdimensionalen Häufigkeitstabelle von Interesse ist. Ein wichtiger Spezialfall der log-linearen Modelle bezieht sich nun auf den Fall einer bestimmten Anzahl unabhängiger Variablen und einer abhängigen Variablen. Dieser Spezialfall einer asymmetrischen Fragestellung, die bereits als LOGIT-MODELL eingeführt wurde, ist danach ein dependenzanalytischer Ansatz, der unmittelbar aus einem bestimmten log-linearen Modell abgeleitet werden kann. Dazu wird eine Variable als abhängig betrachtet. In diesem Fall interessieren uns nur die Beziehungen zwischen dem unabhängigen Merkmalen und dem als anhängig qualifizierten Merkmal. Die Zusammenhangsstruktur der unabhängigen Merkmale ist nicht mehr Gegenstand einer Hypothese, d.h. die Zusammenhänge zwischen den unabhängigen Variablen werden nicht weiter untersucht und als fix angenommen, in dem sie durch das saturierte Submodell für die Prädikatoren dargestellt werden. Man verzichtet somit auf einen Teil der Informationen eines log-linearen Modells, um sich auf die Analyse von Interaktionstermen zu konzentrieren, die die abhängige Variable enthalten. Wir haben das Logit-Modell in dem letzten Abschnitt bereits ausführlich hergeleitet, hier sollen daher nur die Verbindungen zum log-linearen Ansatz behandelt werden. Wir stellen im folgenden kurz die Grundzüge des Logit-Ansatzes in Termen der log-linearen Modellbildung dar und gehen auf die wichtigsten Maßzahlen zur Beurteilung einzelner Effekt-Parameter ein (ausführlich s. z. B. Langeheine 1980, Arminger 1979).

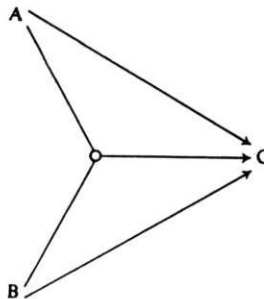
Der Einfachheit halber sein eine dreidimensionale Kontingenztafel mit den unabhängigen Merkmalen A, B und der abhängigen Variablen C angenommen:

Die drei Merkmale haben
die Ausprägungen

A_i , $i = 1, 2$;

B_j , $j = 1, 2$;

C_k , $k = 1, 2$;



Geht man vom saturierten log-linearen Modell aus, so erhält man

$$\text{B.2.1} \quad \log \hat{m}_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC}$$

d.h.

$$\begin{bmatrix} \log \hat{m}_{111} \\ \log \hat{m}_{211} \\ \log \hat{m}_{121} \\ \log \hat{m}_{221} \\ \log \hat{m}_{122} \\ \log \hat{m}_{212} \\ \log \hat{m}_{122} \\ \log \hat{m}_{222} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 \\ 1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & -1 & 1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} \lambda \\ \lambda^A_1 \\ \lambda^B_1 \\ \lambda^C_1 \\ \lambda^{AB}_{11} \\ \lambda^{AC}_{11} \\ \lambda^{BC}_{11} \\ \lambda^{ABC}_{111} \end{bmatrix}$$

wobei im saturierten Modell $\hat{m}_{ijk} = n_{ijk}$ gilt.

Betrachten wir die Merkmale A und B als unabhängige Größen und C als abhängige Variable, so lautet die Grundidee zur Entwicklung des entsprechenden Logit-Modells: Wenn A und B keinen Einfluß auf C haben soll, dann muß das Verhältnis $(\hat{m}_{ij1}/\hat{m}_{ij2})$ für alle (I x J) Ausprägungskombinationen von A und B gleich sein. Das logarithmierte Verhältnis $(\log \hat{m}_{ij1}/\hat{m}_{ij2} = \log \hat{m}_{ij1} - \log \hat{m}_{ij2})$ ist die sogenannte Metrisierung der abhängigen Variablen C, wobei

$$\text{B.2.2} \quad \log \hat{m}_{ij1} - \log \hat{m}_{ij2} = \psi_{ij}^{\bar{C}}$$

$$\text{B.2.2} \quad \log m_{i,-} - \log m_{j,-} = \log m_{i,j}^{TM_j} - \log m_{j,j}^{TM_j}$$

als 'Logit' bezeichnet wird. Beim Logit-Ansatz geht es nun darum, ein Modell für das logarithmierte Verhältnis der konditionalen Häufigkeiten der abhängigen Variablen zu gewinnen - und nicht wie bisher ein Modell für die einzelnen Zellenhäufigkeiten. Aus dem saturierten log-linearen Modell erhalten wir folgende Gleichung des korrespondierenden Logit-Modells durch einfache Subtraktion zweier Modellgleichungen

$$\text{B.2.3 } \psi_{ij}^{\bar{C}} = \log \hat{m}_{ij1} / \hat{m}_{ij2} = \log \hat{m}_{ij1} - \log \hat{m}_{ij2} =$$

$$\begin{aligned} & \lambda + \lambda \frac{A}{i} + \lambda \frac{B}{j} + \lambda \frac{C}{1} + \lambda \frac{AB}{ij} + \lambda \frac{AC}{i1} + \lambda \frac{BC}{j1} + \lambda \frac{ABC}{ij1} - (\lambda + \lambda \frac{A}{i} + \lambda \frac{B}{j} + \lambda \frac{C}{2} \\ & + \lambda \frac{AB}{ij} + \lambda \frac{AC}{i2} + \lambda \frac{BC}{j2} + \lambda \frac{ABC}{ij2}) \\ & = 2 \lambda \frac{C}{1} + 2 \lambda \frac{AC}{i1} + 2 \lambda \frac{BC}{j1} + 2 \lambda \frac{ABC}{ij1} \end{aligned}$$

Alle anderen Parameter entfallen durch die Subtraktion.

Die Vereinfachung in der letzten Zeile der Gleichung B.2.3 folgt daraus, daß bei dichotomen Variablen die Parameter für die beiden Kategorien sich nur durch das Vorzeichen unterscheiden ($\lambda_2^A = -\lambda_1^A$ usw.).

Ersetzt man $2 \lambda = \beta$, so vereinfacht sich die Gleichung zu:

$$\text{B.2.4 } \log (\hat{m}_{ij1} / \hat{m}_{ij2}) = \beta_1^C + \beta_{i1}^{AC} + \beta_{j1}^{BC} + \beta_{ij1}^{ABC}$$

wobei

$$\begin{aligned} \beta_1^C &= 2 \lambda_1^C = \lambda_1^C - \lambda_2^C, \\ \beta_{i1}^{AC} &= 2 \lambda_{i1}^{AC} = \lambda_{i1}^{AC} - \lambda_{i2}^{AC}, \\ \beta_{j1}^{BC} &= 2 \lambda_{j1}^{BC} = \lambda_{j1}^{BC} - \lambda_{j2}^{BC}, \\ \beta_{ij1}^{ABC} &= 2 \lambda_{ij1}^{ABC} = \lambda_{ij1}^{ABC} - \lambda_{ij2}^{ABC}. \end{aligned}$$

Das Modell in Gleichung B.2.4 ist ein Logit-Modell, in dem die logarithmierte Chance, in die Kategorie $C_k = 1$ statt $C_k = 2$ zu fallen, in Abhängigkeit von den Effekten der zwei unabhängigen Variablen A und B dargestellt wird. Mit dem Übergang zum Logit-Modell im Rahmen der log-linearen Modellbildung ist somit ein Übergang zur bedingten Betrachtungsweise verbunden: man interessiert sich für das abhängige Merkmal C unter Vorliegen der Merkmalsausprägungen für A und B. Assoziationen (d.h. die Zusammenhangsstruktur) zwischen den unabhängigen Merkmalen werden nicht betrachtet. Wie die Differenzenbildung (B.2.3.) zeigt, enthält das resultierende Logit-Modell genau dann eine β -Parameter, wenn im zugrundeliegenden log-linearen Modell der entsprechende λ -Parameter mit dem zusätzlichen Index der abhängigen Variablen C enthalten ist.

Logit-Modelle lassen sich - wie in Abschnitt B.1 bereits ausführlich dargestellt - auch unabhängig von hierarchischen log-linearen Modellen betrachten. Unter Annahme der Reparametrisierungsbedingung "centered effects" kann das saturierte Logit-Modell durch folgende Modellgleichung in Matrix-Notation ausgedrückt werden:

$$\begin{bmatrix} \ln(\hat{m}_{111}/\hat{m}_{112}) \\ \ln(\hat{m}_{211}/\hat{m}_{212}) \\ \ln(\hat{m}_{121}/\hat{m}_{122}) \\ \ln(\hat{m}_{221}/\hat{m}_{222}) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_i^A \\ \beta_j^B \\ \beta_{ij}^{AB} \end{bmatrix}$$

bzw.

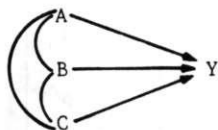
$$\begin{bmatrix} \ln(\hat{m}_{111}/\hat{m}_{112}) \\ \ln(\hat{m}_{211}/\hat{m}_{212}) \\ \ln(\hat{m}_{121}/\hat{m}_{122}) \\ \ln(\hat{m}_{221}/\hat{m}_{222}) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 2 \lambda_1^C \\ 2 \lambda_{11}^{AC} \\ 2 \lambda_{11}^{BC} \\ 2 \lambda_{111}^{ABC} \end{bmatrix}$$

Die Effekte in einem Logit-Modell mit einer dichotomen Zielvariablen sind also nichts anderes als die um die abhängige Variable ergänzten verdoppelten Effekte eines korrespondierenden log-linearen Modells. Logit-Modelle sind somit Spezialfälle log-linearer Modelle und lassen sich daher über letztere testen. Allerdings muß das Modell den mit den fest vergebenen Randsummen korrespondierenden Term (λ_{ij}^{AB}) enthalten.

In der folgenden Abbildung sind mögliche Logit-Modelle für die (asymmetrische) Analyse einer dreidimensionalen Kontingenztafel veranschaulicht. Die Verallgemeinerung auf beliebige höherdimensionale Logit-Modelle ergibt sich völlig analog zur Subtraktion in Gleichung B.2.3, wenn man von den entsprechenden log-linearen Modellen höherer Dimensionen ausgeht.

Beispiel für eine vierdimensionale Tabelle ($A \times B \times C \times Y$) mit Y = abhängige Variable (alle Variablen dichotom):

MODELLHYPOTHESE:



(Haupteffektmodell, d.h. es werden keine Interaktionen zw. den unabh. Var. bzgl. Y angenommen)

ANGEPAßTE RANDVERTEILUNGEN IM LOG-LINEAREN MODELL:

(ABC) (AY) (BY) (CY)

LOGIT-MODELLGLEICHUNG:

$$\log(\hat{m}_{ijk1}/\hat{m}_{ijk2}) = \beta_1^Y + \beta_{i1}^{AY} + \beta_{j1}^{BY} + \beta_{k1}^{CY}$$

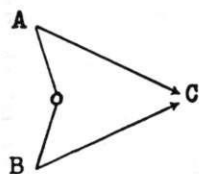
Als Testgröße zur Beurteilung der Modellanpassung verwenden wir wegen seiner statistischen Eigenschaften wieder den Likelihood-Ratio-Test mit der Testgröße CHI_L^2 -abgekürzt X_L^2 :

$$\text{B.2.5} \quad X_L^2 = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K 2n_{ijk} \ln \left(\frac{n_{ijk}}{m_{ijk}} \right)$$

Ist die Modellanpassung unter Annahme der Hypothese ($H_0: \beta^{ABC} = 0$) hinreichend, verbleibt im Rahmen einer Logit-Analyse neben der Parametrisierung des Modells (Berechnung der β -Werte) noch die Aufgabe, die relative Bedeutsamkeit der im Modell spezifizierten Effekte zu bestimmen. Größen, wie etwa das multiple und partielle Bestimmtheitsmaß, die in der Regressions- oder Varianzanalyse Verwendung finden, sind auch für die Logit-Analyse im Rahmen der log-linearen Modellspezifikation verfügbar (vgl. Arminger 1979, S.236f.). Als erstes muß ein geeignetes Vergleichsmodell ("baseline model") definiert werden, dessen CHI_L^2 -Wert als Ausmaß der "Gesamtvariation" in den Daten zu interpretieren ist. Der Begriff "Variation" bezieht sich somit auf Diskrepanzen zwischen unter einem bestimmten Modell erwarteten Häufigkeiten und beobachteten Ausgangsdaten. Das von Arminger (1979) vorgeschlagene "baseline model" geht von der Annahme der paarweisen Unabhängigkeiten der unabhängigen und abhängigen Variablen aus:

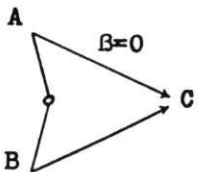
$$\begin{pmatrix} A \\ B \end{pmatrix} \quad C \quad H_1: \{AB\} \{C\} \quad \text{CHI}_L^2(H_1) = \text{Gesamtvariation}$$

log-lineares Modell:



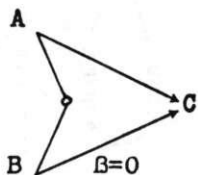
$$\ln \hat{m}_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC}$$

(AB) (AC) (BC)



$$\ln \hat{m}_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{jk}^{BC}$$

(AB) (BC)



$$\ln \hat{m}_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC}$$

(AB) (AC)



C

$$\ln \hat{m}_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB}$$

Logit-Modell:

$$\log \frac{\hat{m}_{ij1}}{\hat{m}_{ij2}} = 2\lambda_1^C + 2 \cdot \lambda_{i1}^{AC} + 2 \cdot \lambda_{j1}^{BC}$$

$$\log \frac{\hat{m}_{ij1}}{\hat{m}_{ij2}} = \beta_1^C + \beta_{i1}^{AC} + \beta_{j1}^{BC}$$

$$\log \frac{\hat{m}_{ij1}}{\hat{m}_{ij2}} = 2\lambda_1^C + 2 \cdot \lambda_{j1}^{BC}$$

$$\log \frac{\hat{m}_{ij1}}{\hat{m}_{ij2}} = \beta_1^C + \beta_{j1}^{BC}$$

$$\log \frac{\hat{m}_{ij1}}{\hat{m}_{ij2}} = 2\lambda_1^C + 2 \cdot \lambda_{i1}^{AC}$$

$$\log \frac{\hat{m}_{ij1}}{\hat{m}_{ij2}} = \beta_1^C + \beta_{i1}^{AC}$$

(Null-Modell)

(β^{AB}, β^{BC}) sei $\chi^2_L(H_1)$ kann da $\chi^2_L(H_1)$ in jedem anderen Modells in Relation gesetzt werden.

$$R^2_{LR}(C \cdot A, B) = (X^2_L\{AB\}\{C\} - X^2_L\{AB\}\{AC\}\{BC\}) / X^2_L\{AB\}\{C\}$$

$$R^2_{LR}$$

Es gibt in diesem Beispiel den % - Anteil R^2_{LR} erklärter Variation von C durch die Prädiktor

$$\begin{aligned} sr^2_{LR}(\beta^A) &= (\text{Fehler ohne A} - \text{Fehler mit A}) / (\text{Fehler ohne A und B}) \\ &= (\chi^2_L\{AB\}\{BC\} - \chi^2_L\{AB\}\{BC\}\{AC\}) / (\chi^2_L\{AB\}\{C\}) \end{aligned}$$

Im Zähler $\beta^A = \beta^B = 0$. - Differenz in der Abweichung zwischen zwei Modellen, die sich in diesem Modell interpretieren. Zur Illustration der Berechnung von Logit-Parametern aus den Effekten des symmetrischen



Die Ergebnisse unterscheiden sich natürlich nicht von jenen der logistischen Regression (vgl. 62.). Die Tabelle zeigt die log-linearen Spezifikationen der jeweiligen Abhängigkeitsmodells.

ERGEBNISSE DER LOG-LINEAREN TABELLENANALYSE					
Hypothetische Abhängigkeitsstruktur	Anzupassende Randverteilungen	χ^2_L	D.F.	p	Erklärte Devianz
Basismodell	(BS) (M)	2856	5	0.0	0.00
$\begin{pmatrix} B \\ S \end{pmatrix} \rightarrow M$	(BS) (MS)	1860	4	0.0	0.35
$\begin{pmatrix} B \\ S \end{pmatrix} \rightarrow M$	(BS) (MB)	98	3	0.0	0.96
$\begin{pmatrix} B \\ S \end{pmatrix} \rightarrow M$	(BS)(MB)(MS)	73	2	0.0	0.97
Saturiertes Modell	(BSM)	0	0	1.0	1.00

Wir wollen im folgenden die LAMBDA-Parameter des Haupteffektmodells näher betrachten:

STADT	MOB		BILDUNG	MOB	
	IMMOBILE	MOBILE		IMMOBILE	MOBILE
Minden	-0,055	0,055	Volksschule	0,487	-0,487
Duisburg	0,055	-0,055	Mitttert	-0,475	0,475
			Sekprima	-0,011	0,011

MOB	
IMMOBILE	MOBILE
-0.022	0,022

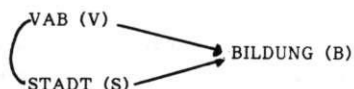
(Wir haben nur die LAMBDA-Werte angegeben, aus denen unmittelbar die LOGIT-Parameter folgen).

Da wir an dem Verhältnis Mobile(Ausprägung 2)/Immobil(e) (Ausprägung 1) interessiert sind, verdoppeln wir jeweils die LAMBDA-Werte der zweiten Kategorie von MOB in den oben tabellierten Parameter-Matrizen. Wir erhalten dann folgende Regressionskoeffizienten aus dem symmetrischen Ansatz:

PARAMETER	log-lineares Modell	Logit-Modell
Stadt(1)	0.055	0.11
Stadt(2)	-0.055	-0.11
Bildung(1)	-0.487	-0.97
Bildung(2)	x 0.475	0.95
Bildung(3)	0.011	0.02
Konstante	0.022	0.04

Die Tabelle enthält exakt die Ergebnisse der logistischen Regression (s.S.60), wenn die Effekt-Kodierung zugrunde gelegt wird.

Die Überprüfung von Logit-Modellen im Rahmen einer log-linearen Tabellenanalyse läßt sich in einfacher Weise auf polytome abhängige Merkmale verallgemeinern. Als Illustrationsbeispiel sei die im Abschnitt (A) betrachtete dreidimensionale Kontingenztafel (STADTxBERUF DES VATERSxBILDUNG) für eine asymmetrische Fragestellung herangezogen:



Ausprägungen:

$V_i = 1, 2, 3;$

$S_j = 1, 2;$

$B_k = 1, 2, 3.$

Die Ergebnisse für eine Sequenz von log-linearen Modellen, in denen jeweils die Interaktionsterme

	Hypothetische Abhängigkeitsstruktur	Anzupassende Randverteilungen	χ^2_L	D.F.:	p:	Erklärte Devianz:
(2)	$\begin{pmatrix} V \\ S \end{pmatrix} \rightarrow B$	(VS) (BV)	1395,3	6	0,0	70,7%
(3)	$\begin{pmatrix} V \\ S \end{pmatrix} \rightarrow B$	(VS) (BS)	3038,7	8	0,0	36,2%
(4)	$\begin{pmatrix} V \\ S \end{pmatrix} \rightarrow B$	(VS) (BV) (BS)	2,1	4	0,7	99,9%
(1)	Basismodell	(VS) (B)	4761,4	10	0,0	----

Hier zeigt sich, daß die Vernachlässigung jedes einzelnen Haupteffektes eine beachtliche Zunahme der Devianz zur Folge hat. Wir akzeptieren daher erst das Modell mit zwei Haupteffekten mit einer erklärten Devianz von nahezu 100%. Der anteilige Devianzzuwachs, der durch Vernachlässigung einer Variablen entsteht, entspricht dem Zuwachs an erklärter Varianz, wenn ein Merkmal zusätzlich in eine Regressionsanalyse aufgenommen wird, d.h. dem relativen Erklärungsbeitrag einer Variablen. Danach kommt der Haupteffekt der Variablen VAB für 63,7% der gesamten Devianz auf und der Haupteffekt der Variablen STADT für 29,2% der Gesamtdevianz auf.

Wenden wir uns nun der Berechnung einzelner Effekt-Parameter zu. Bei der Berücksichtigung einer nominalen abhängigen Variablen mit mehr als zwei Kategorien ergibt sich ein spezielles Problem, das in der Regel unter dem Begriff "multivariate logits" diskutiert wird. Es gibt verschiedene Möglichkeiten, multivariate Logits im Rahmen der log-linearen Modellspezifikation zu definieren (s. Fienberg 1980, Arminger 1982, Arminger 1983), die wir anhand des exemplarischen Beispiels illustrieren wollen. Für den Fall einer abhängigen Variablen mit drei Ausprägungen könne drei Logits gebildet werden:

$$\text{B.2.6} \quad \ln(\hat{m}_{ij1}/\hat{m}_{ij2}) \quad , \quad \ln(\hat{m}_{ij1}/\hat{m}_{ij3}) \quad \text{und} \quad \ln(\hat{m}_{ij2}/\hat{m}_{ij3}) \quad .$$

Ein Logit ist redundant, da

$$\text{B.2.7} \quad (\hat{m}_{ij1}/\hat{m}_{ij2}) \cdot (\hat{m}_{ij2}/\hat{m}_{ij3}) = (\hat{m}_{ij1}/\hat{m}_{ij3}) \quad \text{ist.}$$

Unter Annahme des Haupteffektmodells können dementsprechend zwei unabhängige Kategorienvergleiche als Metrisierung der abhängigen Variablen BILDUNG betrachtet werden. Wir können z.B. jede Kategorie B_k ($k=2,3$) mit der ersten Kategorie B_1 vergleichen, wobei

B_1 = Volksschule

B_2 = Mittelschule bis Tertia

B_3 = Sekunda bis Prima bedeutet.

$$(1) \quad \ln(\hat{m}_{ij1}/\hat{m}_{ij2}) = \beta_0^1 + \beta_i^V(1) + \beta_j^S(1)$$

$$(2) \quad \ln(\hat{m}_{ij1}/\hat{m}_{ij3}) = \beta_0^2 + \beta_i^V(2) + \beta_j^S(2)$$

In der ersten Gleichung betrachten wir die Effekte der unabhängigen Variablen auf das Verhältnis der Wahrscheinlichkeit, einen Volksschulabschluß zu erreichen, zur Wahrscheinlichkeit, einen Abschluß zwischen Mittelschule und Tertia zu erreichen. In der zweiten Gleichung betrachten wir das Verhältnis der Wahrscheinlichkeit, einen Volksschulabschluß zu beobachten, zur Wahrscheinlichkeit, einen Bildungsrang 'Sekunda

- Prima' zu beobachten.

Zur Bestimmung der β -Werte in den zwei Logit-Modellen berechnen wir zunächst die Lambda-Parameter des log-linearen Modells

(SV) (SB) (VB)

wobei uns lediglich die Werte für λ^{SB} , λ^{VB} und λ^B interessieren:

VAB	(λ^{VB})			STADT	(λ^{SB})		
	BILDUNG				BILDUNG		
	Volkssch.	Mittelsch.	SekPrim		Volkssch.	Mittelsch.	SekPrim
BesiBild	-1,715	0,303	1,412	Minden	-0,586	0,383	0,203
MaltMneu	-0,121	-0,099	0,220	Duisburg	0,586	-0,383	-0,203
ArbuntBe	1,836	-0,204	-1,632				

(λ^B)		
BILDUNG		
Volkssch.	Mittelsch.	SekPrim
0,598	0,413	-1,010

Die Logit-Parameter stehen in einfacher Beziehung zu den Lambda-Werten des korrespondierenden log-linearen Modells (SV) (SB) (VB):

Gleichung (1) ("Volksschule" versus Mittelsch. - Tertia):	Gleichung (2) ("Volksschule" versus "Sekunda - Prima"):
$\beta_0^1 = \lambda_1^B - \lambda_2^B = 0,18$	$\beta_0^2 = \lambda_1^B - \lambda_3^B = 1,6$
$\beta_1^V(1) = \lambda_{11}^{VB} - \lambda_{12}^{VB} = -2,04$	$\beta_1^V(2) = \lambda_{11}^{VB} - \lambda_{13}^{VB} = -3,13$
$\beta_2^V(1) = \lambda_{21}^{VB} - \lambda_{22}^{VB} = -0,02$	$\beta_2^V(2) = \lambda_{21}^{VB} - \lambda_{23}^{VB} = -0,34$
$\beta_3^V(1) = \lambda_{31}^{VB} - \lambda_{32}^{VB} = 2,04$	$\beta_3^V(2) = \lambda_{31}^{VB} - \lambda_{33}^{VB} = 3,47$
$\beta_1^S(1) = \lambda_{11}^{SB} - \lambda_{12}^{SB} = -0,97$	$\beta_1^S(2) = \lambda_{11}^{SB} - \lambda_{13}^{SB} = -0,79$
$\beta_2^S(1) = \lambda_{21}^{SB} - \lambda_{22}^{SB} = 0,97$	$\beta_2^S(2) = \lambda_{21}^{SB} - \lambda_{23}^{SB} = 0,79$

(V_1 = Besitz- u. Bildungsbürger; V_2 = Mittelstand alt und neu;
 V_3 = Arbeiter und untere Beamte; S_1 = Minden; S_2 = Duisburg).

Die Werte für die Parameter der dritten (redundanten) Gleichung

$$(3) \quad \ln(\hat{m}_{ij2}/\hat{m}_{ij3}) = \beta_0^3 + \beta_i^V(3) + \beta_j^S(3)$$

erhalten wir mittels der Lambda-Parameter des log-linearen Modells durch die Bildung entsprechender Differenzen:

$$\beta_0^3 = 0,413 - (-1,010) = 1,42$$

$$\beta_1^{V(3)} = 0,303 - 1,412 = -1,11$$

$$\beta_2^{V(3)} = -0,099 - 0,22 = -0,32$$

$$\beta_3^{V(3)} = -0,204 - (-1,632) = 1,43$$

$$\beta_1^{S(3)} = 0,383 - 0,203 = 0,18$$

$$\beta_2^{S(3)} = -0,383 - (-0,203) = -0,18$$

Da aus (B.2.7)

$$\text{B.2.8} \quad \ln(\hat{m}_{ij1}/\hat{m}_{ij2}) + \ln(\hat{m}_{ij2}/\hat{m}_{ij3}) = \ln(\hat{m}_{ij1}/\hat{m}_{ij3}) \quad \text{und}$$

$$\text{B.2.9} \quad \ln(\hat{m}_{ij2}/\hat{m}_{ij3}) = \ln(\hat{m}_{ij1}/\hat{m}_{ij3}) - \ln(\hat{m}_{ij1}/\hat{m}_{ij2}) \quad \text{folgt,}$$

können die Parameter des dritten Logit-Modells auch direkt aus den Werten des ersten und zweiten Logit-Modells berechnet werden:

$$\beta_0^3 = \beta_0^2 - \beta_0^1 ;$$

$$\beta_1^{V(3)} = \beta_1^{V(2)} - \beta_1^{V(1)} ; \quad \beta_2^{V(3)} = \beta_2^{V(2)} - \beta_2^{V(1)} ; \quad \beta_3^{V(3)} = \beta_3^{V(2)} - \beta_3^{V(1)}$$

$$\beta_1^{S(3)} = \beta_1^{S(2)} - \beta_1^{S(1)} ; \quad \beta_2^{S(3)} = \beta_2^{S(2)} - \beta_2^{S(1)} .$$

D.h. die Unterschiede zwischen zwei Kategorien der abhängigen Variablen liegen bzgl. der Effekte der unabhängigen Variablen eindeutig fest, und zwar unabhängig davon, welche Kategorie B₀ als Basiskategorie definiert ist, gegen die wir andere Kategorien der abhängigen Variablen vergleichen. Für k = 1 erhalten wir z.B. den Effekt der Kategorie VAB = 1 bzgl. des Verhältnisses Bildung = 2 versus Bildung = 3 aus folgender (verallgemeinerbaren) Gleichung:

$$\text{B.2.10} \quad \begin{array}{ccccc} \text{Vergleich} & & - & \text{Vergleich} & = & \text{Vergleich} \\ 1 \text{ zu } 3 & & & 1 \text{ zu } 2 & & 2 \text{ zu } 3 \\ \left(\lambda_{11}^{VB} - \lambda_{13}^{VB} \right) - \left(\lambda_{11}^{VB} - \lambda_{12}^{VB} \right) & = & & \left(\lambda_{12}^{VB} - \lambda_{13}^{VB} \right) , \end{array}$$

$$\text{d.h.} \quad -\lambda_{13}^{VB} + \lambda_{12}^{VB} = \lambda_{12}^{VB} - \lambda_{13}^{VB} .$$

Aus Gleichung B.2.10 folgt dann unmittelbar:

$$\text{B.2.11} \quad \beta_1^V (1 \text{ zu } 3) - \beta_1^V (1 \text{ zu } 2) = \beta_1^V (2 \text{ zu } 3)$$

Eine weitere Möglichkeit, multivariate Logits zu bilden, besteht darin, die Basiskategorie ($k=1$) in den Nenner einzelner Logits aufzunehmen. Wir erhalten dann die Logits

$$(1) \quad \ln \frac{\hat{m}_{ij3}}{\hat{m}_{ij1}}; \quad (2) \quad \ln \frac{\hat{m}_{ij2}}{\hat{m}_{ij1}}; \quad (3) \quad \ln \frac{\hat{m}_{ij2}}{\hat{m}_{ij3}}$$

und die Beziehungen

$$\text{B.2.12} \quad \ln (\hat{m}_{ij3}/\hat{m}_{ij1}) + \ln (\hat{m}_{ij2}/\hat{m}_{ij3}) = \ln (\hat{m}_{ij2}/\hat{m}_{ij1})$$

und

$$\text{B.2.13} \quad \ln (\hat{m}_{ij2}/\hat{m}_{ij3}) = \ln (\hat{m}_{ij2}/\hat{m}_{ij1}) - \ln (\hat{m}_{ij3}/\hat{m}_{ij1}).$$

Die Unterschiede zwischen zwei Kategorien der abhängigen Variablen liegen auch in diesem Fall eindeutig fest:

Verhältnis:	2 zu 3	2 zu 1	3 zu 1
	$\beta_1^V (3)$	$\beta_1^V (2)$	$\beta_1^V (1)$
	-1,109	2,018	3,1

Wir geben dieser Form der Logit-Bildung den Vorzug, da in diesem Fall die Parameter für weitere Kategorienvergleiche, in denen die Basiskategorie nicht enthalten ist, ohne die umständliche Berechnungsformel (B.2.10) nach B.2.13 bestimmt werden können. Die Werte für die Parameter $\beta_1^V (2 \text{ zu } 1)$ und $\beta_1^V (3 \text{ zu } 1)$ unterschieden sich von den Parametern $\beta_1^V (1 \text{ zu } 2)$ und $\beta_1^V (1 \text{ zu } 3)$ unter der Annahme der Effektkodierung nur in dem Vorzeichen!

Betrachten wir polytome abhängige Variablen, so können die Ergebnisse sowohl vertikal als auch horizontal gelesen werden. Ein Pfaddiagramm mit multivariaten Logits verdeutlicht im folgenden mögliche Parametervergleiche (s. folgende Seite).

Die Interpretation der Logit-Parameter in vertikaler Lesart ist die folgende:

Wir betrachten zunächst das Verhältnis der Wahrscheinlichkeit, den Schulabschluß "Mittelschule - Tertia" zu erreichen, zur Wahrscheinlichkeit, einen Volksschulabschluß zu erreichen, in Abhängigkeit von den zwei unabhängigen Variablen STADT und VAB. Die Variable BERUF DES VATERS beeinflusst das Verhältnis BILDUNG = "Mittelschule - Tertia" zu BILDUNG = "Volksschule" wie folgt:

Die Ausprägung "Besitz- und Bildungsbürger" erhöhen die Wahrscheinlichkeit, einen mittleren Bildungsabschluß (relativ zur Volksschule) zu erreichen, beträchtlich, während die Herkunftsschicht "Arbeiter und untere Beamte" die Wahrscheinlichkeit erheblich vermindern. Die Ausprägung VAB = "Mittelstand alt und neu" hat keinen

Effekt auf die hier betrachteten Logits. Der Wohnort "Minden" erhöht die Wahrscheinlichkeit für einen mittleren Bildungsrang (relativ zum Volksschulabschluß).

		2 zu 1	3 zu 1	2 zu 3	
VAB	= 1	2,02	3,13	-1,11	
	= 2	0,02	0,34	-0,32	
	= 3	-2,04	-3,47	1,43	
STADT	= 1	0,97	0,79	0,18	
	= 2	-0,97	-0,79	-0,18	
KONSTANTE (GM)		-0,18	-1,6	1,42	

BILDUNG
 2 zu 1
 3 zu 1
 2 zu 3

(VAB₁ = Besitz- und Bildungsbürger; VAB₂ = alter und neuer Mittelstand; VAB₃ = Arbeiter und untere Beamte; STADT₁ = Minden; STADT₂ = Duisburg).

Die numerischen Werte der einzelnen Effekte können zwar größenmäßig verglichen werden, materiell lassen sie sich jedoch nur umständlich interpretieren. Letzteres ist in diesem Anwendungszusammenhang auch nicht so wichtig, da hier nur Größenvergleiche interessant sind, die die relative Bedeutsamkeit einzelner Ausprägungen der unabhängigen Variablen (d.h. die Effektstärken) auf das logarithmierte Verhältnis zweier erwarteter Häufigkeiten abbilden (die Effekte sind analog zu partiellen Regressionskoeffizienten zu interpretieren).

Als Beispiel für eine Vorhersagegleichung im Rahmen eines Logit-Modells berechnen wir die Wahrscheinlichkeit für einen Schulbildungsrang "Mittelschule-Tertia" (relativ zum Volksschulabschluß als Basiskategorie m), wenn

BERUF DES VATERS = Besitz- und Bildungsbürger (V=1) und
 STADT = Minden (S=1)

$$\begin{aligned}
 \ln \left(\frac{\hat{m}_{112}}{\hat{m}_{111}} \right) &= \text{GM}(1) + \beta_1^V(1) + \beta_1^S(1) \\
 &= -0,18 + 2,02 + 0,97 = 2,813
 \end{aligned}$$

Zur Berechnung der Wahrscheinlichkeit (p_2^B) nach Gleichung B.1.10 benötigen wir noch die Werte der Effektparameter aus (3-1) Logit-Gleichungen (eine Gleichung ist, wie wir bereits festgestellt haben, redundant); wir berücksichtigen noch die Effekte der Ausprägungen STADT = 1 und VAB = 1 auf das Verhältnis "Sekunda-Prima" versus "Volksschule":

$$\ln (\hat{m}_{113}^{SVB} / \hat{m}_{111}^{SVB}) = GB(2) + \beta_1^V(2) + \beta_1^S(2) \\ = -1,6 + 3,13 + 0,79 = 2,319$$

Nach der Vorhersagegleichung B.1.10 folgt die Wahrscheinlichkeit (P_2^B) aus

$$P_2^B = \frac{\exp(-0,18 + 2,02 + 0,97)}{1 + \exp(-0,18 + 2,02 + 0,97) + \exp(-1,6 + 3,13 + 0,79)} \\ = \frac{\exp(2,813)}{1 + \exp(2,813) + \exp(2,319)} = \frac{16,66}{27,825} = 0,598$$

d.h. die Wahrscheinlichkeit für den Schulbildungsrang "Mittelschule-Tertia" liegt unter Annahme der Ausprägungskombination VAB = "Besitz- und Bildungsbürger" und STADT = "Minden" bei 0,598. Betrachten wir noch die Vorhersagegleichung für das Verhältnis "Sekunda-Prima" versus "Volksschule" und bestimmen die Wahrscheinlichkeit (P_3^B):

$$P_3^B = \frac{\exp(2,319)}{1 + \exp(2,319) + \exp(2,813)} = 0,365$$

Die Wahrscheinlichkeit (P_1^B) für die Basiskategorie BILDUNG = "Volksschule" folgt nach Gleichung (B.1.10) aus

$$P_1^B = 1 / (1 + \exp(2,319) + \exp(2,813)) = 1/27,82 = 0,0359$$

Die für die Ausprägungskombination VAB=1, STADT=1 im Rahmen der Logit-Analyse vorhergesagten Wahrscheinlichkeiten (P_k^B , $k=1,2,3$) sind nun nichts anderes als die (bedingten) Abstromprozente der Variablen BILDUNG, wenn wir anstelle der beobachteten Zelhäufigkeiten die unter der Modellannahme (VS) (BS) (BV) erwarteten Häufigkeiten zugrunde legen:

STADT	VAB	BILDUNG			
		\hat{m}_{111}	\hat{m}_{112}	\hat{m}_{113}	
		7,044	117,4	71,58	196
1	1	3,59%	59,88%	36,52%	100%

Aus den zwei Logit-Gleichungen mit den Effekten der Ausprägungen VAB= "Besitz- und Bildungsbürger", STADT= "Minden" läßt sich (nach Transformation) direkt der Quotient zweier erwarteter Häufigkeiten (oder bedingter Wahrscheinlichkeiten) berechnen:

$$(1) \quad \hat{m}_{112} / \hat{m}_{111} = \exp(\beta_0^1 + \beta_1^V(1) + \beta_1^S(1)) \\ = \exp(-0,18 + 2,02 + 0,97) = 16,6 \\ = 117,4 / 7,044 = 16,6 \\ \hat{P}_{112} / \hat{P}_{111} = 0,5988 / 0,0359 = 16,6 \quad \text{und}$$

$$\begin{aligned}
 (2) \quad \hat{m}_{113}/\hat{m}_{111} &= \exp(\beta_0^2 + \beta_1^V(2) + \beta_1^S(2)) \\
 &= \exp(-1,6 + 3,13 + 0,79) = 10,16 \\
 &= 71,58/7,044 = 10,16 \\
 \hat{P}_{113}/\hat{P}_{111} &= 0,3652/0,03594 = 10,16 \quad ,
 \end{aligned}$$

d.h. aus den Effekt-Parametern eines multinomialen Logit-Modells mit einer polytomen abhängigen Variablen lassen sich sowohl die einzelnen Kategorienwahrscheinlichkeiten (bzw. Anteilswerte im Sinne von Abstomprozenten) der abhängigen Variablen aus der allgemeinen Vorhersagegleichung des multinomialen Logit-Modells vorhersagen als auch die einzelnen Verhältnisse dieser Wahrscheinlichkeiten mit Hilfe der allgemeinen (k-1) Logit-Gleichungen.

Betrachten wir noch die Logits, die das Verhältnis der Wahrscheinlichkeit "BILDUNG=Sekunda-Prima" zur Wahrscheinlichkeit "BILDUNG=Volksschule" angeben. Wie zu erwarten wirken die Variablen VAB und STADT in gleicher Weise auf B(3) wie auf B(2)-relativ zu B(1). Der Effekt der Kategorie VAB = "alter und neuer Mittelstand" fällt in diesem Fall zwar größer aus, ist aber nach wie vor von geringer Bedeutung. Genau in die andere Richtung wirken die unabhängigen Merkmale, wenn wir das Verhältnis der Kategorien "Mittelschule-Tertia" zu "Sekunda-Prima" betrachten. Die Wahrscheinlichkeit, einen mittleren Bildungsrang relativ zum Abschluß "Sekunda oder Prima" zu erreichen, vermindert sich für Kinder, deren Väter Besitz- oder Bildungsbürger sind, während sie sich erhöht für Kinder, deren Väter Arbeiter oder untere Beamte sind. Die Variable STADT hat auf das Verhältnis B(2) zu B(3) keinen nennenswerten Einfluß mehr.

Wenden wir uns nun der horizontalen Lesart zu. Sie ist dadurch charakterisiert, daß die Effekte einer Kategorie eines unabhängigen Merkmals auf die drei Logits untereinander verglichen werden. Es zeigt sich z.B., daß die Kategorie VAB = "Arbeiter und untere Beamte" generell einen negativen Effekt auf die Wahrscheinlichkeit einer höheren Schulbildung (relativ zur Volksschule) hat; besonders auffallend ist hier der negative Effekt auf das Verhältnis "Sekunda-Prima" versus "Volksschule". Erst mit dem dritten Logit kehrt sich der Trend um: Die Wahrscheinlichkeit für BILDUNG = "Mittelschule bis Tertia" erhöht sich für Kinder unterer Sozialschichten im Vergleich zur Wahrscheinlichkeit für BILDUNG = "Sekunda bis Prima". Betrachten wir einmal die Effekte auf folgende Kategorienvergleiche (Logits) der abhängigen Variablen:

		<u>BILDUNG</u>								
		1	zu	2	1	zu	3	2	zu	3
<u>VAB</u> =	Arbeiter und untere Beamte	+		2,04	+		3,47	+		1,43

(Die Ergebnisse unterscheiden sich nur in dem Wechsel zweier Vorzeichen, da wir nur die Basiskategorie vertauscht haben).

Aus diesen Werten ist ersichtlich, daß die Wahrscheinlichkeit für einen niedrigen Bildungsrang im Vergleich zu einem höheren Bildungsrang für Kinder unterer Sozialschichten generell über dem Gesamtdurchschnitt (GM) liegt.

Betrachten wir abschließend noch die Effekte der Kategorie VAB = "Besitz- und Bildungsbürger". Die durchschnittliche Wahrscheinlichkeit für einen höheren Bildungsrang (relativ zur Volksschule bzw. Mittelschule bis Tertia) liegt bei allen drei Logits in den Subpopulationen VAB = "Besitz- und Bildungsbürger" deutlich über dem jeweils generellen Durchschnitt (GM₁, GM₂ und GM₃).

Die dargestellte Vorgehensweise, die Effekte der unabhängigen Variablen auf eine polytome abhängige Variable simultan mit Hilfe eines multinomialen Logit-Modells zu bestimmen, wird in einem anderen Anwendungszusammenhang auch als "conditional logit modeling" (s. McFadden 1974) oder "multiple logit model" (vgl. Schmidt/Strauss 1975) bezeichnet (''). Im Rahmen dieses Ansatzes können für eine polytome abhängige Variable mit I Ausprägungen allgemein (I-1) Logits z.B. wie folgt definiert werden:

$$\text{B.2.14} \quad \log\left(\frac{\hat{m}_{ijk}}{\hat{m}_{Ijk}}\right), \quad i = 1, 2, \dots, I-1; \quad \log\left(\frac{\hat{m}_{ijk}}{\hat{m}_{\bar{i}-1, jk}}\right), \quad i = 1, 2, \dots, I-1$$

Die beiden Logit-Definitionen unter B.2.14 sind formal gleich, da z.B. für $I = 3$

$$\log\left(\frac{\hat{m}_{1jk}}{\hat{m}_{2jk}}\right) + \log\left(\frac{\hat{m}_{2jk}}{\hat{m}_{3jk}}\right) = \log\left(\frac{\hat{m}_{1jk}}{\hat{m}_{3jk}}\right) \quad \text{gilt.}$$

Wenn wir etwa die Kategorie $I=1$ der abhängigen Variablen als Basiskategorie definieren und Logits bilden, in denen diese Kategorie als Vergleichskategorie dient (d.h. $\hat{m}_{ijk}/\hat{m}_{1jk}$, mit $i = 2, 3$), dann folgen aus den Werten der einzelnen Effekte durch bestimmte Parameterdifferenzen alle weiteren Effekte der unabhängigen Variablen auf die Logits, in denen die erste Kategorie nicht enthalten ist. Die Vorgehensweise, die Parameter eines Logit-Modells simultan durch ein korrespondierendes log-lineares Modell zu schätzen, hat den Vorteil, die Werte der einzelnen Effektparameter für unterschiedliche Logits größtmäßig vergleichen zu können, da sie auf der Basis der geschätzten Häufigkeiten nur eines bestimmten Modells berechnet werden!

Angenommen/anstelle der simultanen Schätzung der Parameter würden wir drei unabhängige, separate Logit-Modelle mit dichotomer abhängiger Variable getrennt mit den jeweils korrespondierenden log-linearen Modelle spezifizieren:

$$(1) \quad \ln \left(\frac{\hat{m}_{ijk}}{\hat{m}_{ijk}} \right) = \beta_0^1 + \beta_i^V(1) + \beta_j^S(1) \quad \text{mit } k = 1, 3$$

$$(2) \quad " = \beta_0^2 + \beta_i^V(2) + \beta_j^S(2) \quad \text{mit } k = 2, 3$$

$$(3) \quad " = \beta_0^3 + \beta_i^V(3) + \beta_j^S(3) \quad \text{mit } k = 1, 2$$

Die resultierenden Parameterwerte aus diesen drei separat geschätzten Logit-Modellen unterscheiden sich u.U. erheblich von denen des konditionalen Logit-Modells. "The point is that, in conditional logit modeling, the proper interpretation of the odds involving any two categories...cannot be made without taking into account information about the other categories" (Swafford 1980, S. 681). Die Information der jeweiligen Restkategorie wird bei den o.g. Definitionen dreier separater Logits ignoriert. Wir halten daher nur folgende alternative Logit-Definitionen im Falle polytomer abhängiger Merkmale für angemessen:

$$B.2.15 \quad \log \left(\frac{\hat{m}_{ijk}}{\sum_{l>i} \hat{m}_{ljk}} \right), \quad i = 1, 2, \dots, I-1;$$

$$B.2.16 \quad \log \left(\frac{\hat{m}_{ijk}}{\sum_l \hat{m}_{ljk}} \right), \quad i = 1, 2, \dots, I-1, \quad \text{wobei } l \neq i,$$

Hat die abhängige, polytome Variable lediglich nominales Skalenniveau, führt die unter B.2.16 angegebene Möglichkeit der Bildung von Logits zu (I-1) abhängigen Dummy-Variablen y_i , wobei

$$y_i = \begin{cases} 1 & \text{wenn BILDUNG den Wert (i) annimmt,} \\ 0 & \text{sonst.} \end{cases}$$

(Eine inhaltlich interessante Kategorie (i) wird somit mit einer Restkategorie verglichen, in der alle Häufigkeiten der übrigen Ausprägungen zusammengefaßt werden). Wenn die Ausprägungen der abhängigen Variablen - wie in unserem Beispiel - geordnet sind, ist es sinnvoll, die sogenannten "continuation ratios" nach B.2.15 zu verwenden (s. Fienberg 1980, S. 86f.). Wir betrachten dann jeweils die Wahrscheinlichkeit, in Kategorie BILDUNG_(i) zu kommen im Verhältnis zur Wahrscheinlichkeit, in eine der Kategorien BILDUNG₍₁₎, ..., BILDUNG_(i) zu fallen. Da die Ausprägungen geordnet sind, interessiert jeweils nur der Vergleich zu den nächst höheren Kategorien. Anstelle eines einzigen multinomialen Logit-Modells treten dann (I-1) Logit-Modelle. Für unser Beispiel erhalten wir somit die Gleichungen

$$(1) \quad \ln (\hat{m}_{ij1} / \hat{m}_{ij2+3}) = \beta_0^1 + \beta_1^V(1) + \beta_j^S(1) \quad \text{mit} \quad n_{ij.} = \sum_{k=1}^3 n_{ijk} \quad \text{und}$$

$$(2) \quad \ln (\hat{m}_{ij2} / \hat{m}_{ij3}) = \beta_0^2 + \beta_1^V(2) + \beta_j^S(2) \quad \text{mit} \quad n_{ij.} = \sum_{k=2}^3 n_{ijk} .$$

Die Parameter der einzelnen Logit-Modelle sind voneinander asymptotisch unabhängig, da die beobachteten, binomialverteilten Häufigkeiten

$$-\frac{n_{ijk}}{\sum_{l \geq i} n_{ljk}} \quad , \quad i = 1, 2, \dots, l-1$$

voneinander asymptotisch unabhängig sind (Fienberg 1980, S.87).

Die Ergebnisse der Schätzung (im Rahmen von zwei log-linearen Modellen) sind auf der folgenden Seite wiedergegeben (Modell I und Modell II). Die Effektparameter der Logit-Modelle ergeben sich in diesem einfachen Fall durch Multiplikation der Werte mit "2". Die Herkunftsschicht "neuer und alter Mittelstand" hat keine Bedeutung sowohl für die Wahrscheinlichkeit, einen Volksschulabschluß zu haben (relativ zu allen höheren Bildungsrängen) als auch für die Wahrscheinlichkeit, einen Bildungsrang "Mittelschule bis Tertia" zu erreichen (relativ zum Abschluß "Sekunda - Prima"). Die negativen Werte $\beta_1^V(1) = 2\lambda_{11}^{VB} = -2,292$ und $\beta_1^V(2) = 2\lambda_{12}^{VB} = -1,11$ bestätigen das bekannte Ergebnis, daß die Kinder von Besitz- und Bildungsbürgern im Durchschnitt eher eine höhere Schulbildung erreichen als Kinder von Arbeitern und unteren Beamten. Die Wahrscheinlichkeit eines Volksschulabschlusses (relativ zu allen höheren Bildungsrängen) vermindert sich schließlich leicht für den Herkunftsort "Minden", d.h. in dieser Stadt wird im Vergleich zu "Duisburg" in dem betrachteten Zeitraum ein höheres Bildungsniveau erreicht.

VAB (V)	BILDUNG		BILDUNG		BILDUNG	
	Volkssch.	Sonstige	Mittelsch.-Tertia	Sekunda-Prima	Sekunda-Prima	Sonstige
Besitzbürger + Bildungsbürger	-1,146	1,146	-0,554	0,554	1,018	-1,018
Mittelstand alt + Mittelstand neu	-0,001	0,001	-0,162	0,162	0,310	-0,310
Arbeiter + untere Beamte	1,147	-1,147	0,716	-0,716	-1,328	1,328
STADT (S)						
Minden	-0,470	0,470	0,089	-0,089	0,157	-0,157
Duisburg	0,470	0,470	-0,089	0,089	-0,157	0,157
KONSTANTE	-0,064	0,064	0,714	-0,714	-1,308	1,308

VS, SB, VB ----	$\chi^2_L: 2,33$		$\chi^2_L: 0,42$		$\chi^2_L: 20,71$	
	d.f.: 2		d.f.: 2		d.f.: 2	
	p : 0,32		p : 0,81		p : 0,00	
	MODELL I		MODELL II		MODELL III	

C. MAGIDSON'S ANSATZ ZUR BESTIMMUNG DER ERKLÄRTEN
VARIANZ EINER NOMINALEN ABHÄNGIGEN VARIABLEN

Bei der Suche nach einem den beobachteten Daten angemessenen Modell beginnt man bei der multivariaten Kreuztabellenanalyse oft mit einem Maximalmodell (saturiertes Modell mit allen Haupt- und Interaktionseffekten) und versucht, es möglichst weit zu vereinfachen, ohne die Devianzen (Abweichungen der erwarteten von den beobachteten Häufigkeiten) zu groß werden zu lassen. Im Gegensatz zur Vorgehensweise im Rahmen log-linearer Modellbildung beginnt man in der klassischen Regressionsanalyse mit einem Minimalmodell, das sukzessive erweitert wird, bis die Prognosefehler nicht mehr wesentlich verringert werden können. Über das Ende dieses Suchprozesses entscheidet üblicherweise ein Hypothesentest, der je nach Suchstrategie eine unterschiedliche Zielsetzung hat: Im Fall log-linearer Modellbildung will man signifikante Abweichungen zwischen Modell und beobachteten Daten vermeiden (χ^2_L -Anpassungstest), während in der Regressionsanalyse alle signifikanten Abweichungen berücksichtigt werden sollen (F-Test).

Unabhängig vom Ausgang dieser Modelltests möchte man die Anpassung des Modells an die Daten durch ein deskriptives Maß beschreiben. Hierzu haben wir in Abschnitt (A) und (B) Analoga zu den multiplen und partiellen Bestimmtheitsmaßen R^2 und sr^2 für den Fall der multivariaten Kontingenztabellenanalyse auf der Basis von normierten χ^2_L -Differenzen aus bestimmten Modellvergleichen diskutiert. Die Anteile erklärter Devianz sind nützliche Hilfen für die Selektion eines möglichst sparsamen Modells und darüberhinaus für die abschließende Beurteilung der relativen Bedeutsamkeit einzelner Effekte sehr anschaulich zu interpretieren.

Maßzahlen der erklärten Devianz für aggregierte Individualdaten in Kontingenztabellen haben jedoch einen Nachteil: Während in der Regressionsanalyse die Gesamtdevianz Auskunft über die Gesamtstreuung der Daten gibt, verwendet man in der Logit-Analyse nach Aggregation der Individualdaten die Abweichungen zwischen erwarteten und beobachteten Häufigkeiten des Basismodells (= Minimalmodell)

$$H_i: \text{Haupt- und Interaktionseffekte gleich Null.}$$

Diese Abweichungen stellen lediglich eine Teildevianz dar, die bei aggregierten Daten immer von der Anzahl der Ausprägungskombinationen der einbezogenen Merkmale abhängt. Die Anwendung log-linearer Modelle auf mehrdimensionale Häufigkeitstabellen gibt daher auch keine Auskunft über die tatsächliche Höhe der erklärten Varianz durch die unabhängigen Merkmale. Durch die Verlagerung der Analyse auf ein gruppenbezogenes Aggregationsniveau ("subclass level") wird gerade diejenige Komponente an der Gesamtvariation ausgeklammert, die in einer "individual-level-analysis" üblicherweise als Fehlervariable eingestuft wird. So werden die für jede Wertekombination der unabhängigen Variablen (Faktorenstufenkombination) gegebenen konditionalen

Verteilungen der abhängigen Variablen auf einzelne, vergleichbare Parameter reduziert, unabhängig davon, wieviele erklärende Merkmale in einem Modell vertreten sind, folgt daraus, daß ein saturiertes Modell immer 100% der Variation des Minimalmodells erklärt. Arminger stellt daher zu Recht fest: "We ... meet with the paradoxical fact that the introduction of additional explanatory variables lowers the explained deviance if we do not introduce all interactions of the old variables with the new one. We miss an analogon to R^2 in regression analysis" (Ariminger 1983, S.1). Wenn wir etwa mit Hilfe des Programms GLIM oder BMDPLR Logit-Modelle auf der Grundlage von Individualdaten berechnen, dann ist die Devianz des Minimalmodells für Individualdaten die richtige Bezugsgröße für die Bestimmung aller multiplen und partiellen Bestimmtheitsmaße, die analog zu denen der Regressionsanalyse interpretiert werden können. In unserer Untersuchung können wir zwar nicht die erklärten Devianzen auf der Basis von Individualdaten berechnen (''), eine Analogon sehen wir jedoch in dem Vorschlag von Magidson (1981), der eine entsprechende Erweiterung des Logit-Modells auf der Grundlage aggregierter Daten darstellt. Wir wollen den Grundgedanken seines Ansatzes daher im folgenden kurz skizzieren (s.a. Langeheine 1980, S. 78ff.).

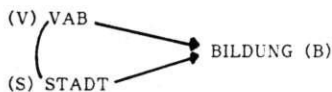
Magidson erweitert zunächst das Konzept der Varianz metrischer Merkmale auf nominale Merkmale. Sei A eine abhängige Variable mit zwei Ausprägungen. Die Varianz läßt sich in diesem einfachen Fall bestimmen nach (Magidson 1981, S. 184):

$$\begin{aligned} \text{C.1} \quad \text{VAR}(A) &= P_1(1 - P_1) \\ &= P_2(1 - P_2) \\ &= P_1 \cdot P_2 \end{aligned}$$

Der allgemeine Ausdruck für die Varianz eines polytomen Merkmals mit J Kategorien folgt dann aus

$$\text{C.2} \quad \text{VAR}(A) = \frac{1}{2} \left(1 - \sum_{j=1}^J P_j^2 \right)$$

In dem gewählten exemplarischen Beispiel



hat die abhängige Variable $J=3$ Ausprägungen, die unabhängige Variable VAB $I_1=3$ und die Variable STADT $I_2=2$ Kategorien. Die Gesamtvariation der abhängigen Variablen läßt sich nun aufspalten in einen erklärten und einen nicht-erklärten Variationsanteil,

analog zur Regressionsanalyse mit einem metrischen abhängigen Merkmal. Aus einer entsprechenden Formel der allgemeinen Variationszerlegung wird der Anteil insgesamt erklärter Varianz (ETA^2) abgeleitet (s. Magidson 1981, S. 189). Anstelle der Darstellung mit Anteilswerten wählen wir im folgenden die Zellenhäufigkeiten der dreidimensionalen Kreuztabelle nach Aggregierung der Individualdaten:

$$\begin{aligned}
 \text{GESAMTDEVIANZ} &= \text{ERKLÄRTE DEVIANZ} + \text{NICHT-ERKLÄRTE DEVIANZ} \\
 (\text{GD}) &= (\text{ED}) + (\text{ND}) \\
 (n_{...} - \sum_j n_{..j}^2 / n_{...}) / 2 &= (\sum_{i_1 i_2 j} I_{i_1 i_2 j}^2 n_{i_1 i_2 j}^2 / n_{i_1 i_2 .} - \sum_j n_{..j}^2 / n_{...}) / 2 \\
 &+ (n_{i_1 i_2 .} - \sum_j I_{i_1 i_2 j} n_{i_1 i_2 j}^2 / n_{i_1 i_2 .}) / 2
 \end{aligned}$$

Aus der Variationszerlegung folgt der multiple Determinationskoeffizient für eine nominale abhängige Variable nach

$$C.3 \quad ETA_{B(VS)}^2 = \frac{(\sum_{i_1 i_2 j} I_{i_1 i_2 j}^2 n_{i_1 i_2 j}^2 / n_{i_1 i_2 .} - \sum_j n_{..j}^2 / n_{...}) / 2}{(n_{...} - \sum_j n_{..j}^2 / n_{...}) / 2} = \frac{ED}{GD}$$

Der Wert beträgt in diesem Beispiel $ETA_{B(VS)}^2 = 0,27$, d.h. es werden insgesamt 27% der Gesamtvariation durch die unabhängigen Merkmale VAB und STADT erklärt.

Das multiple Bestimmtheitsmaß $ETA_{B(VS)}^2$ berücksichtigt nun nicht nur die durch die Einzeleffekte β_{i1}^V ... β_{i2}^S erklärten Varianzanteile, sondern auch den Effekt der Wechselwirkungen beider Merkmale (β^{VS}). Indem wir in der Gleichung C.3 von den empirischen Häufigkeiten $n_{i_1 i_2 j}$ ausgehen, betrachten wir ein saturiertes Modell, das die Interaktionsterme mit einschließt. Den multiplen Determinationskoeffizienten für das Haupteffektmodell erhalten wir, wenn die unter der Modellannahme (VS) (BV) (BS) erwarteten Häufigkeiten $\hat{m}_{i_1 i_2 j}$ in Gleichung C.3 eingesetzt werden:

$$C.4 \quad ETA_{B(V,S)}^2 = \frac{(\sum_{i_1 i_2 j} I_{i_1 i_2 j}^2 \hat{m}_{i_1 i_2 j}^2 / n_{i_1 i_2 .} - \sum_j n_{..j}^2 / n_{...}) / 2}{(n_{...} - \sum_j n_{..j}^2 / n_{...}) / 2}$$

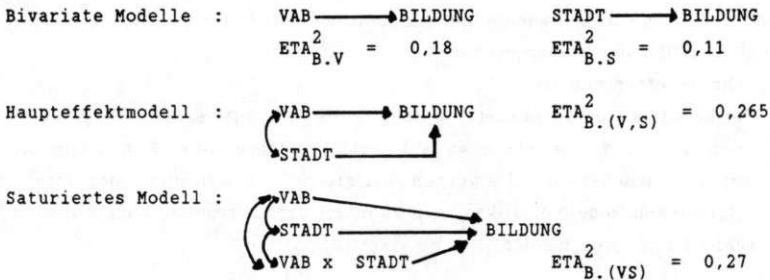
Wir erhalten für das Haupteffektmodell einen erklärten Varianzanteil von 26,5%. Der Wert liegt nur geringfügig unter dem ETA^2 des saturierten Modells.

Da wir in der Regel Zusammenhänge zwischen den unabhängigen Merkmalen beobachten können, tritt das Problem auf, wie (auf unser Beispiel mit zwei unabhängigen Merkmalen bezogen) der erklärte Varianzanteil ($ETA_{B(VS)}^2$) einzelnen Effekten der

Variablen STADT, VAB und dem Interaktionseffekt (VAB x STADT) zugeordnet werden kann, da z.B.

$$\text{ETA}_{B.(S,V)}^2 = \text{ETA}_{B.S}^2 + \text{ETA}_{B.V}^2 \quad \text{gilt.}$$

Neben der globalen Betrachtungsweise interessiert zusätzlich die Evaluierung der relativen Erklärungsanteile der Effekte einzelner Variablen. Die Verallgemeinerung der ETA²-Statistik nach Magidson auf nichtsaturierte Modelle eröffnet nun die Möglichkeit, analog zur Regressionsanalyse² quadrierte semipartielle Korrelationskoeffizienten auf der Grundlage bestimmter ETA²-vergleiche zu berechnen. Die Vorgehensweise sei an unserem Beispiel illustriert:



Die relativen Erklärungsanteile ergeben sich aus den ETA²-Differenzen hierarchisch angeordneter Modelle:

Relativer Erklärungsbeitrag
der Wechselwirkung:

$$\begin{aligned} \text{sr}_{(SV)}^2 &= \text{ETA}_{B.(SV)}^2 - \text{ETA}_{B.(S,V)}^2 \\ &= 0,005 \end{aligned}$$

Relative Erklärungsbeiträge
der Haupteffekte:

$$\begin{aligned} \text{sr}_{(S)}^2 &= \text{ETA}_{B.(S,V)}^2 - \text{ETA}_{B.V}^2 \\ &= 0,085 \end{aligned}$$

$$\begin{aligned} \text{sr}_{(V)}^2 &= \text{ETA}_{B.(S,V)}^2 - \text{ETA}_{B.S}^2 \\ &= 0,155 \end{aligned}$$

Die Vernachlässigung jedes einzelnen Haupteffektes führt zu einer erheblichen Reduktion der erklärten Devianz. Ein Vergleich des Haupteffektmodells mit dem saturierten Modell zeigt, daß die Wechselwirkungen vernachlässigt werden kann, da sie nur eine zusätzliche erklärte Devianz von 0,5% zur Folge hat. Die Ergebnisse machen deutlich, daß die unabhängigen Variablen STADT und VAB erheblich zur Prognose der Bildungsränge beitragen.

Abschließend sei noch kurz auf zwei EDV-Programme eingegangen, die zur Berechnung der ETA²-Statistik nach Magidson benutzt werden können. Eine von R. Langeheine veränderte Version des Programms LOGIT (vgl. Arminger 1979, S. 218ff) liefert im Druckprotokoll neben den Ergebnissen einer Logit-Analyse (z.B. CHI²-Anpassungsstatistiken, Parameter unterschiedlicher Logit-Modellspezifikationen) auch ETA²-Werte für hierarchische Logit-Modellsequenzen.

Quelle: Dr. Rolf Langeheine

Institut für Pädagogik der Naturwissenschaften
Universität Kiel

In dem einführenden Lehrbuch zur multivariaten Analyse qualitativer Daten von R. Langeheine (1980) ist ferner ein kurzes FORTRAN-Programm zur Berechnung von ETA²-Quadrat für log-lineare Modelle auf der Basis der erwarteten Häufigkeiten einer Logit-Modellspezifikation wiedergegeben.

Vorteil: Sehr benutzerfreundlich;

Nachteil: Der Benutzer muß die erwarteten Häufigkeiten für unterschiedliche Modellspezifikationen, wie sie etwa von dem Programm BMDP4F ausgedruckt werden, zunächst einmal eingeben. Bei großen Tabellen und /oder unterschiedlichen Modellspezifikationen kann der daraus resultierende Aufwand schnell sehr groß werden.

D. ZUR REKURSIVEN PFADANALYSE FÜR MEHRDIMENSIONALE KONTINGENZTABELLEN

Die Frage, in welchem Ausmaß einerseits die Ausbildung des Sohnes von der sozialen Herkunft bestimmt ist, andererseits das erreichte Ausbildungsniveau den späteren Beruf determiniert, wird seit Mitte der sechziger Jahre vor allem mit Hilfe pfadanalytischer Modelle beantwortet: Sie wurden von Blau und Duncan (1967) in der klassischen Studie "The American Occupational Structure" für die Untersuchung des Statuszuweisungsprozesses in den USA angewandt. Seit der Veröffentlichung dieser Studie ist das Modell der Pfadanalyse das weithin gebräuchliche Verfahren für die Mobilitätsforschung in verschiedenen Gesellschaften geworden (s. u.a. Müller 1975; Halsey et al. 1980). Da wir ursprünglich auch in unserer Untersuchung von pfadanalytischen Modellen ausführlichen Gebrauch machen wollten, soll im folgende das pfadanalytische Grundmodell kurz dargestellt werden, wobei dann insbesondere die Grenzen der Anwendung auf nominale Merkmale interessieren.

Ausgangspunkt der Pfadanalyse ist eine Menge von Kausalhypothesen über die Beziehungsstruktur zwischen Merkmalen. Die Analyse beginnt mit der Überführung der Hypothesen in eine graphisches Pfadmodell, dem sogenannten Pfaddiagramm. Für die

Konstruktion eines Pfaddiagramms liegen folgende Konventionen vor (vgl. Opp/Schmidt 1976):

Es werden drei Arten von Variablen unterschieden. Vorgegebene oder unabhängige, nicht in dem Modell erklärte Variablen werden als "Exogene" eingeführt. Von anderen hypothetisch abhängige Variablen sind "Endogene". Es wird angenommen, daß sich die Variation einer endogenen Variablen aus einer Linearkombination von kausal vorgelagerten Variablen erklären läßt. Bei jeder hypothetischen Beziehung wird die kausal vorgelagerte Variable als Determinierende und die kausal nachgelagerte Variable als Determinierte bezeichnet. Da sich die endogenen Variablen nicht vollständig durch kausal vorgelagerte Merkmale erklären lassen, werden so genannte Residualvariablen eingeführt. Sie repräsentieren alle Einflußfaktoren, die nicht explizit in das Modell mit aufgenommen werden konnten. Die ursprünglich von Wright (1934) konzipierte Pfadanalyse ist nur auf rekursive Modelle anwendbar, d.h. es werden keine zirkulären Kausalbeziehungen zugelassen. Neuere Ansätze lassen zwar auch nichtrekursive Modellstrukturen zu, wir wollen uns jedoch auf die Darstellung von rekursiven Modellen beschränken.

Der Schlüssel zu einem Kausalmodell besteht in einem Pfaddiagramm, in dem die Abhängigkeitsbeziehungen zwischen den Variablen eines Systems spezifiziert werden. Für die Darstellung des Pfaddiagramms gilt:

- Hypothetisch postulierte kausale Beziehungen werden durch Pfeile von determinierenden zu determinierten Variablen dargestellt.
- Postulierte korrelative Beziehungen (d.h. ungerichtete Zusammenhänge) zwischen exogenen Variablen werden durch gekrümmte Doppelpfeile dargestellt.
- Residualvariablen werden durch Pfeile auf endogene Variablen symbolisiert.
- Die Stärke kausaler Beziehungen wird durch Pfadkoeffizienten (p_{ji}) gekennzeichnet, wobei i die determinierte Variable und j die determinierende Variable bezeichnet.

Die Konstruktion eines einfachen Pfadmodells soll an folgendem einfachen Beispiel erläutert werden:

Endogene Variablen:

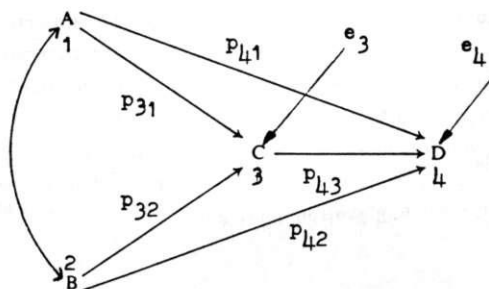
(C) = Bildungsrang des Sohnes;

(D) = Beruf des Sohnes mit 30 Jahren;

Exogene Variablen:

(A) = Herkunftsschicht;

(B) = Zeit.



Ein solches Diagramm läßt sich in eine Reihe von Strukturgleichungen übersetzen, wobei für jede endogene Variable eine Gleichung aufzustellen ist. Das resultierende lineare Gleichungssystem ist die mathematische Abbildung des Hypothesensystems und damit die Grundlage zur quantitativen Schätzung der unbekannten Pfadkoeffizienten; hier besteht es aus zwei Gleichungen:

$$\text{I.} \quad C = p_{31}A + p_{32}B + p_{3e3}e_3 ;$$

$$\text{II.} \quad D = p_{41}A + p_{42}B + p_{43}C + p_{4e4}e_4$$

Die Pfadkoeffizienten können mit Hilfe der OLS-Methode ("ordinary least Squares method") geschätzt werden. Wird von unstandardisierten Variablen ausgegangen, entspricht der standardisierte partielle Regressionskoeffizient (β_{ij}) dem pfadkoeffizienten (p_{ij}). Indirekte Effekte ergeben sich aus der Multiplikation von direkten Effekten:

Indirekter Effekt der Variablen A	Indirekter Effekt der Variablen B
$p_{31} \times p_{43}$	$p_{32} \times p_{43}$

Die Summe der direkten und indirekten Effekte ergibt jeweils den totalen Kausaleffekt einer Variablen:

<u>VARIABLE:</u>	<u>TOTALER KAUSALEFFEKT:</u>	<u>BZGL. D. ABH. VARIABLEN:</u>
C	p_{43}	D
A	$p_{41} + (p_{31} \times p_{43})$	D
B	$p_{42} + (p_{32} \times p_{43})$	D
A	p_{31}	C
B	p_{32}	C

Werden a priori bestimmte Pfadkoeffizienten gleich Null angenommen (z.B. $p_{41} = 0$, zwischen den Variablen "Herkunftsschicht" und "Beruf des Sohnes" wird explizit keine Beziehung vermutet), dann liegen mehr Schätzgleichungen als zu schätzende Parameter vor und das Modell ist überidentifiziert. Für diesen Fall werden unterschiedliche Methoden vorgeschlagen, um das postulierte Pfadmodell mit Null-Pfaden auf Anpassung an die beobachtete Datenstruktur zu prüfen (ausführlich s. Opp/Schmidt 1976).

Die Anwendung der Pfadanalyse erfordert zunächst intervallskalierte Variablen; exogene nominale Variablen können auch durch die Überführung der Dummy-Variablen als quasi-metrische Merkmale in eine Regressionsgleichung aufgenommen werden. Drei grundlegende Voraussetzungen müssen zusätzlich erfüllt sein:

- (1) Die kausale Reihung der Variablen muß eindeutig bestimmbar sein. In der Regel kann die zeitliche Ordnung in der Folge der verschiedenen Variablen die Richtung der kausalen Abhängigkeiten festlegen.

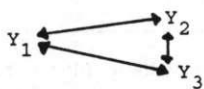
- (2) Das postulierte Kausalmodell kann als vollständig betrachtet werden, d.h. es sind alle relevanten Einflußgrößen in dem Kausalschema spezifiziert. Da die Annahme eines geschlossenen Systems problematisch ist, wird in Pfadmodellen für jede endogene Variable die Residualvariable definiert, die alle ungemessenen oder unbekannten Variablen außerhalb des Systems repräsentieren, die einen Einfluß auf die endogenen Variablen haben könnten.
- (3) Die Residualvariablen dürfen einander nicht korrelieren. Diese Annahme ist etwa in unserem Beispiel gleichbedeutend mit der Forderung, daß es keine im Modell nicht spezifizierten Variablen gibt, die einen Einfluß auf den Bildungsrang des Sohnes und den Beruf des Sohnes haben.

Da die Voraussetzung intervallskalierter endogener Variablen in unserer Untersuchung nicht erfüllt ist, stellt sich die Frage, ob in Analogie zur Pfadanalyse bei quantitativen Daten auch simultane Gleichungssysteme für diskrete Daten geschätzt und anhand beobachteter Daten überprüft werden können. Goodman (1973) hat einen Lösungsansatz auf der Grundlage des log-linearen Wahrscheinlichkeitsmodells vorgeschlagen und die Analogie der Prozedur zur Testung von Kausalhypothesen bei diskreten Daten zur Pfadanalyse mit metrischen Variablen betont (''). Wir wollen im folgenden die grundlegende Idee der rekursiven Pfadanalyse von Kontingenztabellen aufgreifen und weiterführende Ansätze, die vor allem in der Ökonometrie behandelt werden, vorstellen und deren Reichweite für unsere Fragestellung diskutieren.

Unterstellen wir zunächst ein symmetrisches log-lineares Modell der Form $(Y, Y, | Y)$ zwischen den abhängigen und den unabhängigen Variablen Y_i mit den Kategorien $i=1, -1$. Das log-lineare Modell unter der Annahme $(H_0: \lambda_{ijk}^{Y_1 Y_2 Y_3} = 0)$ lautet dann:

$$\lambda_{ijk}^{Y_1 Y_2 Y_3} = 0$$

Hypothetisches Modell: Log-lineare Modellgleichung:



$$\{Y_1 Y_2\} \{Y_1 Y_3\} \{Y_2 Y_3\}$$

$$\ln(\hat{p}_{ijk}) = \lambda^0 + \lambda_{i1}^{Y_1} + \lambda_{j2}^{Y_2} + \lambda_{k3}^{Y_3} + \lambda_{ij}^{Y_1 Y_2} + \lambda_{ik}^{Y_1 Y_3} + \lambda_{jk}^{Y_2 Y_3}$$

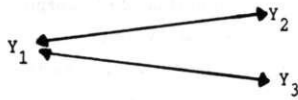
Zwei Logit Modelle mit Y_i und y_i als abhängige Variablen können aus dem allgemeinen log-linearen Modell abgeleitet werden:

$$L_{2|13} = \ln \frac{P(Y_2=+1 | Y_1 Y_3)}{P(Y_2=-1 | Y_1 Y_3)} = 2\lambda^{Y_2} + 2\lambda^{Y_2 Y_1} + 2\lambda^{Y_2 Y_3} \\ = \beta_0^2 + \beta_1^2 Y_1 + \beta_3^2 Y_3 \quad \text{und}$$

$$L_3 | 12 = \ln \frac{P(Y_3=+1 | Y_1 Y_2)}{P(Y_3=-1 | Y_1 Y_2)} = 2 \lambda^{Y_3} + 2 \lambda^{Y_3 Y_1} + 2 \lambda^{Y_3 Y_2} \\ = \beta_0^3 + \beta_1^3 Y_1 + \beta_2^3 Y_2,$$

wobei die Restriktion $2 \lambda^{Y_2 Y_3} = 2 \lambda^{Y_3 Y_2}$ gilt, d.h. $\beta_3^2 Y_3 = \beta_2^3 Y_2$.

Die beiden Gleichungen des multinomialen Logit-Modells korrespondieren direkt mit nur einem log-linearen Modell in die Maximum-Likelihood-Schätzungen für die Parameter der Logit-Modelle sind gleich dem zweifachen bestimmter ML-Schätzungen des gemeinsamen log-linearen Modells. Schmidt/Strauß (1975) bezeichnen diesen Typ der Schätzung simultaner Gleichungssysteme als "simultaneous logit model". Die separat gewonnenen Schätzwerte der gemeinsamen Parameter zweier Gleichungen (hier: $\lambda^{Y_2 Y_3}$ bzw. $\lambda^{Y_3 Y_2}$) sind sowohl konsistent als effizient und exakt gleich, d.h. $\beta_3^2 Y_3 = \beta_2^3 Y_2$. Betrachten wir nun einmal das hypothetische Modell:



Log-lineares Modell $(Y_1 Y_2) (Y_1 Y_3)$:

$$\ln(\hat{P}_{ijk}) = \lambda + \lambda^{Y_1} + \lambda^{Y_2} + \lambda^{Y_3} + \lambda^{Y_1 Y_2} + \lambda^{Y_1 Y_3}$$

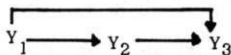
Konditionale log-lineare Modelle:	Anzupassende Randverteilungen im log-linearen Modell:
$L_1 23 = \beta_0^1 + \beta_2^1 Y_2 + \beta_3^1 Y_3$	$(Y_2 Y_3) (Y_1 Y_2) (Y_1 Y_3)$
$L_2 13 = \beta_0^2 + \beta_1^2 Y_1$	$(Y_1 Y_3) (Y_2 Y_1)$
$L_3 12 = \beta_0^3 + \beta_1^3 Y_1$	$(Y_1 Y_2) (Y_3 Y_1)$

Die letzten zwei Gleichungen der konditionalen log-linearen Modelle führen zu effizienten Schätzungen, die direkt aus den Schätzungen des gemeinsamen log-linearen Modells $(Y_1 Y_2) (Y_1 Y_3)$ folgen. Die Parameterschätzungen aus der ersten Gleichung sind weder effizient noch korrespondieren sie mit Parametern des log-linearen Modells $(Y_1 Y_2) (Y_1 Y_3)$ oder mit den Interaktionstermen der zweiten bzw. dritten Gleichung. "Thus whether or not the different CLM equations have to be estimated jointly to get efficient estimates will depend on whether they correspond to different LLM models or to the same LLM model" (Maddala 1982, S. 108).

Die Ableitung konditionaler log-linearer Modelle aus einem gemeinsamen log-linearen Modell zeigt, daß hier lediglich ein spezieller Typ von Zusammenhangsanalyse vorliegt, der mit Strukturgleichungsmodellen analog zur Pfadanalyse wenig gemeinsam hat, sieht

man einmal von der graphischen Veranschaulichung einer komplexen Abgangigkeitsstruktur ab. Ein alternatives Konzept ist in dem sogenannten rekursiven Logit-Modell zu sehen, das hier kurz dargestellt werden soll (ausfuhrlich s. Fienberg 1980, S. 94ff; Lee 1981, S. 56ff; Maddala 1982, S. 108ff).

Wir gehen von folgendem hypothetischen rekursiven Pfadmodell aus, in dem die Abhangigkeitsbeziehungen zwischen drei dichotomen Merkmalen spezifiziert sind:



Die statistische Analyse hat die Marginalverteilung der exogenen Variablen Y_j .. beruckichtigen, die bedingte Verteilung der endogenen Variablen Y_i (gegeben Y_j) und die bedingte Verteilung der endogenen Variablen Y_i (gegeben Y_j und Y_k). Aus dem nominalen Skalenniveau der drei Merkmale folgen als relevante Verteilungen die Wahrscheinlichkeitsverteilungen

$$P(Y_1) \times P(Y_2 | Y_1) \times P(Y_3 | Y_1, Y_2) = P(Y_1, Y_2, Y_3)$$

Jede Wahrscheinlichkeitskomponente entspricht einem Logit Modell:

$$(1) \quad L_1 = \ln \frac{P(Y_1 = +1)}{P(Y_1 = -1)} = 2 \lambda_1^1 ;$$

$$(2) \quad L_{2|1} = \ln \frac{P(Y_2 = +1 | Y_1)}{P(Y_2 = -1 | Y_1)} = 2 \lambda_2^2 + 2 \lambda_{12}^2 Y_1$$

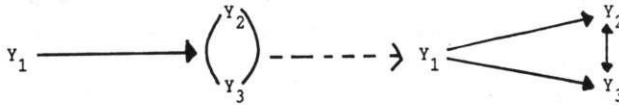
$$(3) \quad L_{3|12} = \ln \frac{P(Y_3 = +1 | Y_1, Y_2)}{P(Y_3 = -1 | Y_1, Y_2)} = 2 \lambda_3^3 + 2 \lambda_{13}^3 Y_1 + 2 \lambda_{23}^3 Y_2 + 2 \lambda_{123}^3 Y_1 Y_2$$

mit den anzupassenden Randverteilungen des jeweils korrespondierenden log-linearen Modells:

(1) (Y_1)	(2) $(Y_1 Y_2)$	(3) $(Y_1 Y_2 Y_3)$
= Randverteilung	= zweidimensionale	= dreidimensionale
von Y_1	Tabelle $Y_1 \times Y_2$	Tabelle $Y_1 \times Y_2 \times Y_3$

Jede Modellgleichung wird in diesem Fall also separat mit Hilfe eines bestimmten log-linearen Modells geschatzt. Da jeweils ein multinomiales Logit-Modell spezifiziert wird, mu auch jeweils die gemeinsame Hufigkeitsverteilung der als unabhangig betrachteten Merkmale in dem log-linearen Modell angepat werden. Auf eine Besonderheit in diesem Zusammenhang sei noch hingewiesen: Der Wert fur den Interaktionsterm λ_{12}^3 der dreidimensionalen Tabelle (Y_1, Y_2, Y_3) ist nicht identisch mit dem Wert des gleichen Terms λ_{12}^2 in der zweidimensionalen Tabelle, da die Dimensionierung der Ausgangstabelle unmittelbar die Ergebnisse der Schatzung beeinflust. Das bislang diskutierte Modell kann als vollstandig rekursiv bezeichnet werden (vgl. Maddala 1982, S. 110). Wir konnen auch ein partiell rekursives Modell spezifizieren, in dem die exogene

Variable Y_1 , die endogenen Variablen Y_2 und Y_3 , gemeinsam beeinflusst:



Statistisches Modell: $\bar{P}(\bar{Y}_1, \bar{Y}_2, \bar{Y}_3) = \bar{P}(\bar{Y}_1) \times \bar{P}(\bar{Y}_2, \bar{Y}_3 | \bar{Y}_1)$

Die mit dieser Modellspezifikation korrespondierenden konditionalen log-linearen Modelle

$$(1) \quad \ln \frac{P(Y_2 = +1 | Y_1 Y_3)}{P(Y_2 = -1 | Y_1 Y_3)} = 2\lambda_{02}^2 + 2\lambda_{23}^2 Y_3 + 2\lambda_{21}^2 Y_1 \quad \text{und}$$

$$(2) \quad \ln \frac{P(Y_3 = +1 | Y_1 Y_2)}{P(Y_3 = -1 | Y_1 Y_2)} = 2\lambda_{03}^3 + 2\lambda_{32}^3 Y_2 + 2\lambda_{31}^3 Y_1,$$

wobei wiederum $2\lambda_{23}^2 Y_3 = 2\lambda_{32}^3 Y_2$ gilt,

können auch in die vereinfachte log-lineare Form

$$\ln P(Y_2, Y_3 | Y_1) = \lambda + \lambda^{Y_2} + \lambda^{Y_3} + \lambda^{Y_2 Y_3} + \lambda^{Y_2 Y_1} + \lambda^{Y_3 Y_1} + \lambda^{Y_1 Y_2 Y_3}$$

überführt werden. Auch das partiell rekursive Modell kann im Gegensatz zu einem Strukturgleichungsmodell als Spezialfall einer Zusammenhangsanalyse angesehen werden.

Der Grund ist in der eingeschränkten strukturellen Interpretationsfähigkeit des Modellansatzes zur Analyse von Beziehungen zwischen den endogenen Variablen des Systems zu sehen. Interaktionen zwischen Y_1 und Y_2 in der log-linearen Modellgleichung können darauf zurückzuführen sein, daß die Residuen der Einzelgleichungen korreliert sind, weil für beide Gleichungen wichtige exogene Variablen nicht in den Ansatz einbezogen wurden. Sie können andererseits darin begründet sein, daß ein rekursives oder ein simultanes System vorliegt. Zwischen diesen Alternativen kann statistisch nicht immer eindeutig diskriminiert werden (vgl. Lee 1981, S. 56ff; Maddala 1982, S. 122ff). Heckmann (1978) hat schließlich daraufhingewiesen, daß die Parametrisierung dieses Modells nicht reich genug ist, um strukturelle Zusammenhänge von rein statistischen Assoziationen zu unterscheiden, wie es etwa im Rahmen von Strukturgleichungsmodellen für metrische Merkmale möglich ist. Die vollständig rekursive Spezifikation eines Modells, die zu einer hierarchischen Sequenz von Logit-Modellen führt, scheint derzeit aufgrund der schätztechnischen Vorteile und für den logischen Ablauf von mehrstufigen Kausalmodellen eine angemessene Vorgehensweise zu sein.

Zwei wesentliche Unterschiede gegenüber einer Pfadanalyse mit metrischen Variablen sind jedoch hervorzuheben:

- (1) In dem multinomialen Logit-Modell gehen wir von einem multiplikativen Ansatz aus, in dem die bedingten Erwartungswerte der abhängigen Variablen $E(Y/X)$ auf eine nichtlineare Regressionsfunktion, z.B.

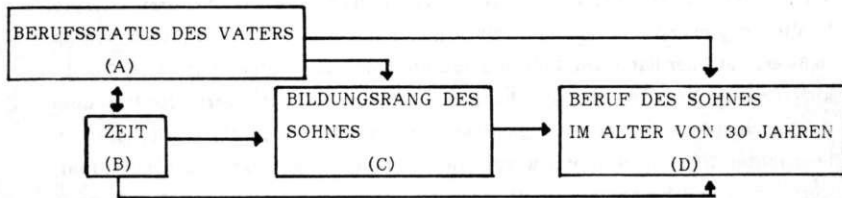
$$E(Y_k | X) = \frac{\exp(\sum_{j=1}^{m-1} \beta_{kj} X_j)}{1 + \sum_{j=1}^{m-1} \exp(\sum_{j=1}^{m-1} \beta_{kj} X_j)} \quad (k=1\dots m), \text{ folgen.}$$

Damit wird die Berechnung von indirekten Effekten unmöglich, die in der Pfadanalyse mit metrischen endogenen Variablen dem Produkt standardisierter Pfadkoeffizienten entsprechen (ausführlich s. Arminger 1983, S. 18ff). Betrachten wir noch einmal das Modell des Statuszuweisungsprozesses (s. das erste Beispiel in diesem Abschnitt). In den Strukturgleichungen I. und II. schätzen wir die Meßwerte auf der Basis von Individualdaten in den abhängigen Variablen C und D aufgrund der Meßwerte der unabhängigen Variablen. Die Meßwerte der Variablen C sind insofern invariant gegenüber dem Status dieses Merkmals, d.h. die beobachteten Individualwerte werden in Gleichung I. als Funktion der Individualwerte in A und B dargestellt, die gleichen Werte werden auch in Gleichung II. zur Vorhersage von Meßwerten in D herangezogen. Im multinomialen Logit-Modell betrachten wir dagegen konditionale Logits als Metrisierung einer nominalen abhängigen Variablen: Der Status der Variablen A und B hat einen Einfluß auf die Logits von C, und der Status von C in der zweiten Gleichung (und nicht die Logits) beeinflußt die Logits von D. Die abhängige Variable C in der ersten Gleichung eines Logit-Modells ist nicht identisch mit ihrem Status als unabhängige Variable in dem zweiten Logit-Modell. Ein weiteres Problem ist darin zu sehen, daß die Analyse einer Sequenz von Logit-Modellen auf der Grundlage unterschiedlich dimensionierter Häufigkeitstabellen beruht; die Anzahl der Zeilen der Design-Matrix \underline{X} (d.h. die Ausprägungskombinationen der jeweils betrachteten unabhängigen Merkmale) variiert demnach mit der Zahl der simultan betrachteten Merkmale. Damit ist es nicht mehr möglich, die Parameterwerte von Logit-Modellen mit unterschiedlicher Merkmalszahl zu vergleichen oder gar multiplikativ zu indirekten Effekten zu verknüpfen (s.a. Goodman 1973, S. 1181).

- (2) Der zweite wichtige Unterschied ist in der jeweiligen Definition der Fehlervariation zu sehen: in der Pfadanalyse mit metrischen endogenen Variablen wird die Fehlervariation ($s_y^2 - \hat{\sigma}^2 = 1 - MR^2$) einer Residualvariablen (ϵ_i) zugeordnet, die den Einfluß aller nicht spezifizierten Variablen auf die jeweilige endogene Variable zusammenfaßt. Damit wird in einem Pfaddiagramm zugleich explizit deutlich gemacht, wie hoch der Erklärungsanteil der verschiedenen Teilsysteme (d.h. Strukturgleichungen bzw. Regressionen) ist. Im Logit-Ansatz betrachten wir keine Modelle mit Individualdaten, sondern gehen zu gruppierten Einheiten in

einer Häufigkeitstabelle über, den Subpopulationen (=Ausprägungskombinationen der unabhängigen Merkmale). Der Grad der Modellanpassung wird nun nicht mehr als Anteil erklärter Varianz gefaßt, sondern als Differenz zwischen den unter einer Modellannahme theoretisch erwarteten Häufigkeiten und beobachteten Werten berechnet. Die so ermittelte Devianz (CHI-Quadrat-Wert) ist nicht modellexternen Einflußgrößen zuzurechnen, sondern der Vernachlässigung von Interaktionseffekten höherer Ordnung (ein saturiertes Modell reproduziert die bedingte Verteilung eines abhängigen Merkmals perfekt, d.h. die Devianz ist in diesem Fall gleich Null).

Zum Schluß dieses Abschnitts wollen wir kurz auf das Modell des Statuszuweisungsprozesses eingehen, in dem wir alle vier Merkmale als nominalskaliert annehmen:



Zur Überprüfung der Verträglichkeit dieses Systems mit den beobachteten Daten sind zwei separate Logit-Modelle zu schätzen bzw. zwei Strukturgleichungen aufzustellende eine für die Merkmale C und D, auf die Pfeile zugehen. Nehmen wir zunächst an, daß die Daten jeweils durch ein Haupteffektmodell hinreichend reproduziert werden und gehen wir ferner von der log-linearen Spezifikation eines Logit-Modells aus. Den beiden Strukturgleichungen entsprechen dann folgende Logit-Modelle:

$$(1) \quad \text{Logit}_{ij}^{AB|C} = \ln \frac{n_{ij}^{AB|C}}{n_{ij}^{AB|C}} = \beta_0^k + \beta_i^{(A)} + \beta_j^{(B)}, \text{ mit } k \neq 1$$

(log-lineares-Modell: {AB} {AC} {BC}) ;

$$(2) \quad \text{Logit}_{ijk}^{ABC|D} = \ln \frac{n_{ijk}^{ABC|D}}{n_{ijk}^{ABC|D}} = \beta_0^{111} + \beta_i^{(A)} + \beta_j^{(B)} + \beta_k^{(C)}, 1 \neq 1$$

(log-lineares Modell: {ABC} {AD} {BD} {CD}) .

Zur Überprüfung der ersten Gleichung wird die dreidimensionale Tabelle (AxBxC) herangezogen, die zweite Gleichung wird anhand der vollständigen (AxBxCxD)-Tabelle geprüft. Das Gesamtsystem wird in zwei Komponenten zerlegt, wobei für jede Komponente eine Hierarchie von unterschiedlichen Modellannahmen auf Verträglichkeit mit

den beobachteten Daten geprüft werden kann:

Tabelle	Hypothese	Angepaßte Randverteilungen
(CBA)	H1	(C), (AB)
	H2	(AC), (BC), (AB)
	H3 ₁	(AC), (AB)
	H3 ₂	(BC), (AB)
	saturiertes Modell	(CAB)
(CBAD)	H4	(D), (ABC)
	H5	(CD), (BD), (AD), (ABC)
	H6 ₁	(CD), (BD), (ABC)
	H6 ₂	(CD), (AD), (ABC)
	H6 ₃	(BD), (AD), (ABC)
	saturiertes Modell	(CBAD)

(Den Hypothesen innerhalb eines Subsystems entsprechen jeweils unterschiedliche Null-Pfade in dem Pfaddiagramm).

Ist für jedes Subsystem ein den Daten hinreichend angepaßtes Modell qualifiziert, kann das Gesamtsystem anhand der Testgröße $\chi^2_{L(H^*)}$, die durch Summation der einzelnen $\chi^2_{L^*}$ -Werte sowie der Freiheitsgrade definiert ist, auf Verträglichkeit mit den beobachteten Daten beurteilt werden (vgl. Goodman 1973; Fienberg 1980).

Unsere abschließende Bemerkung zur rekursiven Pfadanalyse bei diskreten Daten betrifft die Parametrisierung des Modells, d.h. die Berechnung der Parameter einzelner Logit-Modelle des Gesamtsystems: Hat eine Variable mehr als zwei Ausprägungen, so ist der entsprechende Pfad durch einen Parametervektor gekennzeichnet; liegen mehrere polytome Variablen vor, sind ganze Parametermatrizen in das Pfaddiagramm einzutragen. Ein übersichtliche Darstellung der einzelnen Effekte in Form eines Pfaddiagramms ist dann nicht mehr möglich. Die Zuordnung von numerischen Werten zu Pfeilen eines Pfaddiagramms empfiehlt sich daher u.E. nur für den Fall dichotomer endogener Variablen.

Eine mögliche Lösung dieses Problems sehen wir in der zusammenfassenden Charakterisierung der Effektstärke eines unabhängigen Merkmals durch relative Erklärungsanteile (sowohl auf der Grundlage der $\chi^2_{L^*}$ -Statistik als auch nach Magidson's Ansatz zur Berechnung der ETA'-Statistik)(¹¹).

**E. PARTIELLE ASSOZIATIONSMAßE ZUR INFORMATIONSREDUKTION
DER PARAMETERSCHÄTZUNGEN IN LOG-LINEAREN
WAHRSCHEINLICHKEITSMODELLEN**

Als Folge des kategorialen Charakters aller Daten in unserer Untersuchung, d.h. auch der endogenen Variablen BILDUNG und BERUF DES SOHNES, ist der Störterm in der Schätzgleichung diskret und die üblich Normalverteilungsannahme der Regressionsanalyse nicht mehr angemessen. In dieser Forschungssituation ist die Anwendung von bedingten log-linearen Modellen sinnvoll, mit deren Hilfe die Einflüsse der Kategorien der unabhängigen Variablen auf die Wahrscheinlichkeiten für das Auftreten der Kategorien der abhängigen Variablen analysiert werden können. Die bereits in Abschnitt (A) diskutierte Schätzmethode basiert auf der u.a. von Birch (1963) und Goodman (1978) entwickelten Modellklasse der multivariaten log-linearen Wahrscheinlichkeitsmodelle, die um die Klasse der bedingten Logit-Modelle erweitert wurden.

Log-lineare Modelle wurden vor allem von Goodman (1972) zur Untersuchung von kausalen Beziehungen herangezogen. Für den Fall dichotomer Variablen kann die Kausalstruktur zwischen kategorialen Merkmalen in Pfaddiagrammen analogen Schaubildern zusammengefaßt werden. Falls die Variablen aber nicht dichotom sind, führt dies - wie Kawasaki/Zimmermann (1981) zu Recht anführen - zu relativ komplexen Schaubildern. Die Besonderheit eines Pfadmodells für polytome Kategorialvariablen besteht gerade darin, daß auf den Pfeilen, die von oder zu einer nominalen Variablen verlaufen, keine Einzelwerte mehr als Pfadkoeffizienten geschrieben werden können, sondern Matrizen der Haupteffekte. Die Pfaddiagrammdarstellung verliert dann schnell jeglichen Nutzen, da das Beziehungsgeflecht (ausgedrückt in den Parameterschätzwerten eines Logit-Modells) kaum noch zu überschauen ist.

Beide Autoren schlagen daher ein Verfahren vor, das die Informationen der Parameterschätzungen mittels bekannter statistischer Assoziationsmaße reduziert (Kawasaki/Zimmermann 1981, S. 93ff.). In diesem Abschnitt wollen wir die einzelnen Stufen der Modellkomprimierung darstellen und anhand des bereits verwendeten exemplarischen Beispiels illustrieren.

Gegeben sei eine zweidimensionale Kontingenztafel ($A \times B$):

A \ B	B			
	B ₁	B ₂	B ₃	
A ₁	n ₁₁	n ₁₂	n ₁₃	n _{1.}
A ₂	n ₂₁	n ₂₂	n ₂₃	n _{2.}
	n _{.1}	n _{.2}	n _{.3}	n _{..}

In dem saturierten Modell kann die bedingte Wahrscheinlichkeit, daß ein Individuum in die j -te Kategorie des Merkmals B fällt, wenn es sich in der i -ten Kategorie des Merkmals A befindet, nach Logarithmierung als Funktion der Haupt- und Interaktionseffekte ausgedrückt werden:

$$\log P(i_A = 1, j_B = 1) = \mu + \alpha_A(1) + \alpha_B(1) + \beta_{AB}(1,1)$$

$$\log P(i_A = 1, j_B = 2) = \mu + \alpha_A(1) + \alpha_B(2) + \beta_{AB}(1,2)$$

$$\log P(i_A = 1, j_B = 3) = \mu + \alpha_A(1) + \alpha_B(3) + \beta_{AB}(1,3)$$

$$\log P(i_A = 2, j_B = 1) = \mu + \alpha_A(2) + \alpha_B(1) + \beta_{AB}(2,1)$$

$$\log P(i_A = 2, j_B = 2) = \mu + \alpha_A(2) + \alpha_B(2) + \beta_{AB}(2,2)$$

$$\log P(i_A = 2, j_B = 3) = \mu + \alpha_A(2) + \alpha_B(3) + \beta_{AB}(2,3)$$

$$\sum_{i_A} \sum_{j_B} P_{i_A j_B} = 1$$

Das Gleichungssystem kann unter Annahme der Reparametrisierungsbedingung der Effekt-Kodierung

$$(1) \quad \sum_{i=1}^I \alpha_i^A = \sum_{j=1}^J \alpha_j^B = 0 \quad \text{und} \quad \sum_{ij} \beta_{ij}^{AB} = 0$$

in die vereinfachte log-lineare Form

$$(2) \quad \underline{\log P_{AB}} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 0 & -1 & 0 \\ 1 & -1 & 0 & 1 & 0 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} \mu \\ \alpha_A(1) \\ \alpha_B(1) \\ \alpha_B(2) \\ \beta_{AB}(1,1) \\ \beta_{AB}(1,2) \end{bmatrix}$$

überführt werden, wobei μ die Konstante darstellt. Die Parameter α^A und α^B sind die Haupteffekte des Modells, die Parameter β_{ij}^{AB} quantifizieren die von Zeilen- und Spalteneffekten unabhängige Größe der Interaktion zwischen der Kategorie A_i und B_j .

Der Vektor $\log P_{AB}$ enthält die bedingten Wahrscheinlichkeiten p_{ij} als Werte der abhängigen Variablen.

Gleichung (2) beschreibt detailliert die Beziehung zwischen den Kategorien der zwei Variablen. Da wir hier an der Analyse des Gesamtzusammenhangs interessiert sind, verwenden wir ein von Kawasaki/Zimmermann (1981) vorgeschlagenes Verfahren, das die Informationen der Parameterschätzungen (α_i^A , α_j^B und β_{ij}^{AB}) mittels bekannter statistischer Assoziationsmaße reduziert. Diesem Ansatz liegt die Überlegung zugrunde, daß z.B. Gleichung (2) in delogarithmierter Form als Produkt von komponentenweisen Wahrscheinlichkeiten geschrieben werden kann (vgl. Kawasaki/Zimmermann 1981, S. 93):

$$\begin{aligned}
 (3) \quad P_{ij} &= \exp\{\mu + \alpha_A(i) + \alpha_B(j) + \beta_{AB}(i, j)\} \\
 &= \frac{\{ \sum_i \exp(\alpha_A(i)) \} \cdot \{ \sum_j \exp(\alpha_B(j)) \} \cdot \{ \sum_i \sum_j \exp(\beta_{AB}(i, j)) \}}{\sum_i \sum_j \exp\{\alpha_A(i) + \alpha_B(j) + \beta_{AB}(i, j)\}} \\
 &= \frac{\exp(\alpha_A(i))}{\sum_i \exp(\alpha_A(i))} \cdot \frac{\exp(\alpha_B(j))}{\sum_j \exp(\alpha_B(j))} \cdot \frac{\exp(\beta_{AB}(i, j))}{\sum_i \sum_j \exp(\beta_{AB}(i, j))} \\
 &= C \cdot p_i^{\bar{A}} \cdot p_j^{\bar{B}} \cdot p_{ij}^{\bar{AB}}
 \end{aligned}$$

mit

$$\begin{aligned}
 p_i^{\bar{A}} &= \frac{\exp(\alpha_A(i))}{\sum_i \exp(\alpha_A(i))} \\
 p_j^{\bar{B}} &= \frac{\exp(\alpha_B(j))}{\sum_j \exp(\alpha_B(j))} \\
 p_{ij}^{\bar{AB}} &= \frac{\exp(\beta_{AB}(i, j))}{\sum_i \sum_j \exp(\beta_{AB}(i, j))}
 \end{aligned}$$

und

$$C = \frac{\{ \sum_i \exp(\alpha_A(i)) \} \cdot \{ \sum_j \exp(\alpha_B(j)) \} \cdot \{ \sum_i \sum_j \exp(\beta_{AB}(i, j)) \}}{\sum_i \sum_j \exp\{\alpha_A(i) + \alpha_B(j) + \beta_{AB}(i, j)\}}$$

("exp" bezeichnet den Delogarithmus).

In Gleichung (3) ist C eine Konstante und $p_i^{\bar{A}}$, $p_j^{\bar{B}}$ und $p_{ij}^{\bar{AB}}$ sind die komponentweisen Wahrscheinlichkeiten, wobei die Beziehung

$$p_{ij}^{\bar{A}\bar{B}} = \frac{\exp(\beta_{AB}(i, j))}{\sum_i \sum_j \exp(\beta_{AB}(i, j))}$$

als von Zeilen- und Spalteneffekten unabhängige, partielle Kontingenztafel zwischen den Merkmalen A und B angesehen und mit Assoziationsmaßen ausgewertet werden kann. Aus Gleichung (3) wird deutlich, daß für den Fall

$$\alpha_A(i) = 0 \quad \text{und} \quad \alpha_B(j) = 0 \quad (\text{für alle } i \text{ und } j)$$

(4) $p_{ij} = p_{ij}^{\bar{A}\bar{B}}$ gilt.

"Therefore, the component probability represents a pure or partial probability in the sense that the influence of all other configurations is simultaneously removed" (Kawasaki/Zimmermann 1981, S. 94).

Um die Bedeutung der komponentweisen Wahrscheinlichkeiten besser verstehen zu können, sei einmal das Modell der statistischen Unabhängigkeit in der zweidimensionalen Tabelle (A x B) angenommen. Unter der Hypothese $\beta_{ij}^{AB} = 0$ (für alle i und j) lautete die Modellgleichung in log-linearer Form:

$$(5) \quad \log p_{ij} = \mu + \alpha_i^A + \alpha_j^B,$$

$$\text{bzw.:} \quad p_{ij} = \exp(\mu) \cdot \exp(\alpha_i^A) \cdot \exp(\alpha_j^B) = p_{i.} \cdot p_{.j}$$

Gleichung (5) impliziert nun, daß die Komponentwahrscheinlichkeiten $p_{ij}^{\bar{A}\bar{B}}$ mit den Marginalwahrscheinlichkeiten von A und B exakt übereinstimmen. Da in diesem Fall

$$c \cdot p_{ij}^{\bar{A}\bar{B}} = 1 \quad \text{für alle } i \text{ und } j \text{ gilt,}$$

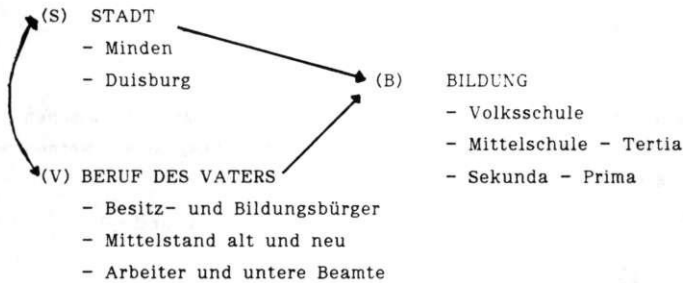
nimmt die Gleichung (3) folgende reduzierte Form an:

$$(6) \quad p_{ij} = p_i^{\bar{A}} \cdot p_j^{\bar{B}}.$$

Gleichung (6) entspricht der klassischen Formulierung des Unabhängigkeitsmodells im zweidimensionalen Fall (s. Gl. 5). Die Komponentwahrscheinlichkeiten $p_{ij}^{\bar{A}\bar{B}}$ ($i=1, \dots, I$ und $j=1, \dots, J$) sind dann gleich einer Konstanten:

$$p_{ij}^{\bar{A}\bar{B}} = \frac{\exp(\beta_{ij}^{AB})}{\sum_{i=1}^I \sum_{j=1}^J \exp(\beta_{ij}^{AB})} = \frac{1}{I \cdot J} \quad \text{für alle } i \text{ und } j.$$

Für die empirische Anwendung betrachten wir noch einmal das exemplarische Beispiel



Das korrespondierende Logit-Modell, das der Hypothese

"0 : Keine Interaktion zwischen STADT und VAB im Hinblick auf die abhängige Variable BILDUNG

entspricht, kann als Produkt folgender komponentweisen Wahrscheinlichkeiten geschrieben werden:

$$\Pr(B / V, S) = C \cdot \overline{\Pr(B)} \cdot \overline{\Pr(V)} \cdot \overline{\Pr(S)} \cdot \overline{\Pr(V, S)} \cdot \overline{\Pr(B, S)} \cdot \overline{\Pr(B, V)}$$

wobei wir die Ausdrücke $\overline{\Pr(B, S)}$ und $\overline{\Pr(B, V)}$ als partielle Kontingenztabellen zwischen den Variablen S B und V B ansehen.

Unter Annahme des log-linearen Modells (SV) (SB) (VB) ermitteln wir zunächst die Werte für die Interaktionseffekte β_{ik}^{SB} und β_{jk}^{VB} .

		BILDUNG			Delogarithmierte Parameterschätzwerte:		
		= 1	= 2	= 3			
VAB	= 1	-1,715	0,303	1,412	0,18	1,35	4,10
	= 2	-0,121	-0,099	0,220	0,88	0,90	1,25
	= 3	1,836	-0,204	-1,632	6,27	0,81	0,19

		BILDUNG			Delogarithmierte Parameterschätzwerte:		
		= 1	= 2	= 3			
STADT	= 1	-0,586	0,383	0,203	0,56	1,47	1,22
	= 2	0,586	-0,383	-0,203	1,79	0,68	0,82

Nach Gleichung (3) können die komponentweisen Wahrscheinlichkeiten $\overline{\Pr(V, B)}$ und $\overline{\Pr(S, B)}$ aus den delogarithmierten Parameterschätzwerten berechnet werden:

		BILDUNG		
		= 1	= 2	= 3
VAB	= 1	0,01	0,08	0,26
	= 2	0,06	0,06	0,08
	= 3	0,39	0,05	0,01

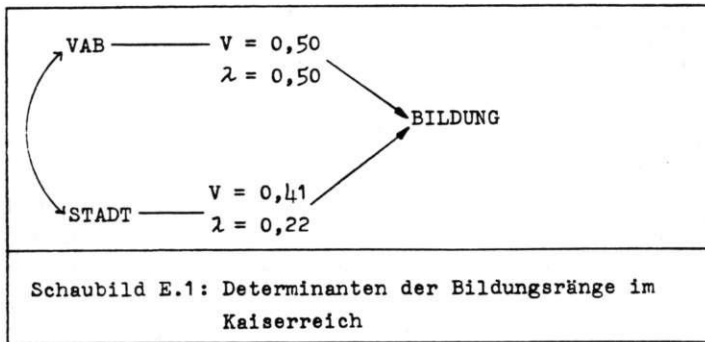
		BILDUNG		
		= 1	= 2	= 3
STADT = 1		0,08	0,22	0,19
= 2		0,27	0,10	0,12

Für die Auswertung dieser zwei partiellen Kontingenztabellen wollen wir die Assoziationsmaße LAMBDA und CRAMER's V heranziehen (''):

- Das LAMBDA (X) wird hier in seiner asymmetrischen Form verwendet und liegt im Wertebereich immer zwischen 0 und 1. Die Maßzahl gibt an, in welchem Grad uns die Kenntnis der Klassifikation einer unabhängigen Variablen zur Vorhersage der Klassifikation einer abhängigen Variablen hilft. Die LAMBDA-Werte repräsentieren folglich die proportionale Fehlerreduktion, die bei der Vorhersage der einen Variablen durch die Auswertung der Informationen einer anderen Variablen erzielt wird (PRE-Kriterium, d.h. "proportional reduction of error" - Konzept). Es ist ein Maß, das keine Restriktion der Tabellengröße kennt und eine anschauliche Interpretation hat.
- CRAMER's V ist ein auf den χ^2 - Wert basierendes Zusammenhangsmaß, das unabhängig von der Tabellengröße immer zwischen 0 und 1 variieren kann. Daher eignet sich diese? Muß insbesondere für den Vergleich von Kontingenztabellen mit unterschiedlicher Zeilen- bzw. Spaltenzahl. Ein Wert von 0 tritt dann auf, wenn die betrachteten Variablen vollständig voneinander unabhängig sind.

Die Ergebnisse der Informationskomprimierung auf der Grundlage der berechneten komponentweisen Wahrscheinlichkeiten in den partiellen Kontingenztabellen $V \rightarrow B$ und $S \rightarrow B$ sind in dem Schaubild auf der folgenden Seite zusammengefaßt.

Die Schätzergebnisse bezüglich der Assoziation zwischen den beteiligten Variablen machen deutlich, daß zwischen den unabhängigen Merkmalen VAB und STADT und dem abhängigen Merkmal BILDUNG ein jeweils starker (partieller) Zusammenhang besteht. Bezüglich der LAMBDA-Werte zeigen sich allerdings Unterschiede in dem Grad der Vorhersage der Kategorien der abhängigen Variablen BILDUNG: Aufgrund der Kenntnis der beruflichen Stellung des Vaters wird man im Durchschnitt 50% weniger Fehlklassifikationen bei der Vorhersage der Variablen BILDUNG (unter Kontrolle des unabhängigen Merkmals STADT) begehen, während man bei Kenntnis der Variablen STADT nur eine relative Fehlerreduktion von 22% erreicht. Daraus folgt, daß man bei Kenntnis des Vaterberufs - relativ zur Kenntnis der Stadtzugehörigkeit - den Bildungsrang genauer vorhersagen kann.



F. DIE ANALYSE DISKRETER DATEN IM RAHMEN VERALLGEMEINERTER MODELLE

Multivariate Analyseverfahren für metrische abhängige Merkmale wie Varianz-, Kovarianz- und Regressionsanalyse, die man unter dem Oberbegriff des allgemeinen linearen Modells (ALM als formal-statistische Rahmentheorie) subsumieren kann (vgl. z.B. Holm (1979), Rochel (1983)), gehören heute zum üblichen Instrumentarium der empirischen Sozialforschung. Die wesentlichen Ziele der Anwendung klassischer multivariater Analyseverfahren sind Vorhersage bzw. Erklärung metrischer Beobachtungsdaten durch ein Set unabhängiger Merkmale.

Erklärung und Prognose setzen allgemein die Beantwortung von zwei Fragen voraus:

1. Was wird erklärt (Definition der abhängigen Variablen)?
2. Welche Merkmale werden zur Erklärung herangezogen (Auswahl der unabhängigen Variablen)?

Für die Modellspezifikation gilt, daß zwischen dem angenommenen Zusammenhang zwischen Prädikatorvariablen und Zielvariablen ein mathematisches Modell bzw. eine mathematische Funktion tritt, wobei Vorhersagefehler nicht auszuschließen sind. Es sind daher zwei weitere Fragen zu beantworten:

3. Wie werden Zielvariable und Prädiktoren miteinander verknüpft (Annahme eines Abhängigkeitsmodells)?
4. Welche Fehlerstruktur wird angenommen?

Die Aussagen über die einzelnen Größen der vier Grundfragen bilden die gemeinsame Struktur von Modellen zur Erklärung von Beobachtungsdaten. Während das ALM in der empirischen Sozialforschung nahezu ausschließlich für Problemstellungen mit metrischen abhängigen Variablen definiert und diskutiert wurde, ist ein formal-statistischer Ansatz wenig beachtet worden, der die o.g. gemeinsame Struktur von Erklärungsmodellen als

Ausgangspunkt für eine Ausweitung des ALM auf kategoriale Kriteriumsvariablen hat. Dieser Ansatz ist von Neider und Wedderburn (1972) unter dem Begriff "verallgemeinerte Modelle" entwickelt und in das EDV-Programm GLIM ("general linear models") umgesetzt worden.

Seit ihrer Einführung haben verallgemeinerte lineare Modelle wachsende Bedeutung in Theorie und Anwendung erlangt, was sich in der zunehmenden Anzahl von Arbeiten zu diesem Thema zeigt (s. z.B. Arminger (1982), Arminger (1983), McCullagh/Nelder (1983), Andreß (1986)). Das formal-statistische Konzept bietet einen gemeinsamen Rahmen u.a. für Modelle mit metrischen abhängigen Variablen, für log-lineare und logistische Modelle.

Im folgenden stellen wir zunächst die allgemeine Struktur verallgemeinerter Modelle dar und diskutieren anschließend anhand eines exemplarischen Beispiels die Anwendung auf nominale Merkmale.

Verallgemeinerte Modelle werden durch drei Komponenten charakterisiert:

- (1) Die stochastische Natur des Fehlerterms:

Die Verteilung der Fehlerterme und damit der abhängigen Variablen entstammt der sogenannten Exponentialfamilie.

- (2) Der lineare Prädiktor:

Die beobachteten unabhängigen Variablen bilden einen linearen Prädiktor.

- (3) Die Verknüpfungsfunktion:

Es existiert eine linearisierende Verknüpfungsfunktion ("link function"), die als Bindeglied zwischen den Erwartungswert der abhängigen Variablen und dem linearen Prädiktor fungiert.

Wir wollen zunächst die Struktur verallgemeinerter Modelle am Beispiel der multiplen Regressionsanalyse illustrieren.

Die Modellgleichung lautet:

$$F.1 \quad Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + e_i, \quad i=1, \dots, n$$

oder in Matrix-Form

$$F.2 \quad \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \dots & x_{k1} \\ 1 & x_{12} & \dots & x_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \dots & x_{kn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} e_1 \\ e_1 \\ \vdots \\ e_n \end{bmatrix}$$

$$\underline{Y} = \underline{X} \underline{\beta} + \underline{e}$$

$$(n \times 1) \quad (n \times k+1) (k+1 \times 1) \quad (n \times 1)$$

Formal gesehen werden die n Meßwerte in der Zielvariablen Y (d.h. y_i , $i=1, \dots, n$) als lineare Funktion von k unabhängigen Einflußgrößen und einer Konstanten dargestellt. Für jede unabhängige Variable x_j ($j=1, \dots, k$) liegen die Variablenwerte X_{ij} vor. Die Einflußgewichte β_j ($j=2, \dots, k$) sind für die Prädikatoren x_j spezifisch und bilden neben der Konstanten die unbekannten Parameter des Modells. Da wir die abhängige Variable nicht perfekt als gewichtete Summe von Einflußgrößen erklären können, sind die Fehler e_i ($i=1, \dots, n$) als Zufallsgröße in die Modellgleichung aufzunehmen. Für die Fehlerkomponente gibt es eine Reihe möglicher Ursachen: Das Modell kann z.B. wesentliche erklärende Variablen nicht enthalten oder in dem spezifizierten Modell sind nur unbedeutende Prädikatoren berücksichtigt. Schließlich kann auch die Annahme eines linearen Ansatzes unangemessen sein.

Im allgemeinen linearen Modell (kurz: ALM) wird der y -Vektor durch einen systematischen Anteil ($\underline{X}\underline{\beta}$) und einen unsystematischen Anteil (\underline{e}) dargestellt. Beide Modellteile werden schließlich additiv zu $\underline{y} = \underline{X}\underline{\beta} + \underline{e}$ verknüpft.

Schätzt man die Werte von β_j mit der Methode der kleinsten Quadrate (OLS), dann lassen sich folgende Kennwerte berechnen (wir verwenden zur kompakteren Darstellung die Matrixnotation):

KENNWERTE DES ALLGEMEINEN LINEAREN MODELLS	
$\underline{y} = (y_i)$	$i = 1, \dots, n$
$\underline{x} = (x_i)$	$i = 1, \dots, n$
$\underline{\beta} = (\beta_j)$	$j = 1, \dots, k+1$
$\underline{b} = (b_j)$	$j = 1, \dots, k+1$ (Schätzwerte für $\underline{\beta}$)
Lösungsgleichung:	Schätzwerte:
$\underline{b} = (\underline{X}'\underline{X})^{-1} \underline{X}'\underline{y}$	$\hat{\underline{y}} = \underline{X}\underline{b}$
$s^2 = \frac{1}{(n-p)} \sum_{i=1}^n (y_i - \hat{y}_i)^2$	
(geschätzte Fehlervarianz)	
$R^2 = 1 - \left(\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \right)$	
(multiples Bestimmtheitsmaß)	

Das multiple Bestimmtheitsmaß (R^2) ist ein Maß für die Güte der Modellanpassung, d.h. für den Grad der Prognosequalität. Der Wert wird als Anteil erklärter Varianz interpretiert.

Wenn die unabhängigen Variablen nominal/ordinal skaliert sind, können sie nach

Transformation in 0/1-kodierte Dummy-Variablen in die Modellgleichung des ALM aufgenommen werden.

Die Einführung von Stellvertretervariablen für die Ausprägungen diskreter unabhängiger Merkmale in die Regressionsgleichung führt zu den Spezialfällen des ALM:

- ein- oder mehrfaktorielle Varianzanalyse (die unabhängigen Variablen weisen lediglich nominales/ordinales Skalenniveau auf);
- Kovarianzanalyse (sowohl metrische Prädikatoren als auch Dummy-Variablen sind in der Regressionsanalyse spezifiziert).

Damit sind bereits zwei Spezialfälle des ALM genannt, wenn die abhängige Variable metrisches Skalenniveau aufweist.

In dem klassischen Modell der Regressionsanalyse gelten noch folgende Annahmen:

- (1) $E(e_i) = 0;$
- (2) $Var(e_i) = \sigma^2;$
- (3) $Cov(e_i, e_j) = 0$ für $i, j = 1, \dots, n$ und $i \neq j$.

Mit anderen Worten: Der Erwartungswert für den Fehlerterm ist gleich Null, es gilt die Annahme, daß die Fehlerterme für alle Elemente die gleiche Varianz aufweisen (Homoskedastizität), aufgrund der Unabhängigkeit der Stichprobenelemente sind die Fehler unkorreliert.

Die Regressionsgleichung F.1 bzw. F.2 führt zu der üblichen Zerlegung der y_i -werte in einen durch das Modell erklärten Anteil (systematische Komponente, dargestellt in der Form des Erwartungswertes $E(Y_i) = \mu_i$) ... inen Rest (die Fehlerkomponente e_i). Wir können die Regressionsgleichung mit p Größen (Konstante + k Prädikatoren, $k+1=p$) daher auch wie folgt ausdrücken:

$$F.3 \quad E(Y_i) = \mu_i = \sum_{j=1}^p \beta_j x_{ij}; \quad i = 1, \dots, N,$$

bzw. in Matrixform:

$$F.4 \quad E(Y) = \mu \quad \text{mit} \quad \mu = X\beta.$$

Um den Übergang zum verallgemeinerten linearen Modell (im folgenden mit GLM - General Linear Model - abgekürzt) zu veranschaulichen, zerlegen wir die Gleichung des Regressionsmodells zunächst in drei Komponenten:

- (1) Zufallskomponente:

Es liegen N voneinander unabhängige Beobachtungswerte (y_i) der abhängigen Variablen Y mit konstanter Varianz und dem Erwartungswert $E(Y_i) = \mu_i$ vor.

- (2) Systematische Komponente:

Die gewichtete Summe (d.h. die Linearkombination) der unabhängigen Variablen

bildet den linearen Prädiktor:

$$F.5 \quad \eta = \sum_1^p \beta_j x_j,$$

(3) Verbindungsfunktion:

Die dritte, wesentliche Komponente des GLM-Ansatzes hat die Aufgabe, den Modellteil

$$F.6 \quad y_i = \mu_i + e_i, \quad i=1, \dots, N$$

mit dem linearen Prädiktor F.5 zu verknüpfen. Diese Aufgabe fällt der sog. "link function" zu, welche die Elemente des linearen Prädiktors auf die Elemente von U umkehrbar eindeutig abbildet:

$$F.7 \quad \eta_i = g(\mu_i) + e_i, \quad i=1, \dots, N$$

so daß sich das Modell (F.6) entsprechend umformulieren läßt:

$$F.8 \quad Y_i = g^{-1}(\eta_i) + e_i, \quad i=1, \dots, N$$

Um eine Übereinstimmung zwischen ALM und GLM deutlich zu machen, muß zunächst eine spezielle Link-Funktion genannt werden, die Identitätsfunktion

$$\mu = \eta$$

Unter Annahme der Identitätsfunktion als Spezialfall geht das GLM in das ALM mit metrischer Zielvariable über, d.h. wir erhalten das klassische Regressionsmodell mit normal verteilten und voneinander unabhängigen Fehlertermen:

$$Y_i \sim N(\mu_i, \sigma^2), \quad \mu = \eta, \quad \eta = \sum_1^p \beta_j x_j.$$

Aussagen dieses speziellen Abhängigkeitsmodells:

(1) Definition der abhängigen Variablen:

Beobachtungswerte einer metrischen Zielvariablen

(2) Definition des linearen Prädiktors:

- nur metrische Variablen: multiple Regression
- nur Dummy-Variablen : Varianzanalyse
- metrische und Dummy-V. : Kovarianzanalyse

(3) Verbindungsfunktion: Identitätsfunktion

(4) Fehlerstruktur:

Die Fehlerkomponente in dem klassischen linearen Modell mit metrischer Zielvariable ist durch die Normalverteilung definiert

Eine für die Parameterschätzung und die inferenzstatistische Absicherung der Ergebnisse wichtige Entscheidung betrifft die Annahme einer geeigneten Fehlerstruktur. Je nach zu analysierenden abhängigen Merkmal (z.B. metrische, dichotome Zielvariable oder Häufigkeiten einer Kontingenztafel) ist eine bestimmte Fehlerverteilung zu wählen. In der Statistik gibt es eine Reihe theoretischer Verteilungen, von denen wir die drei wichtigsten im folgenden kurz charakterisieren wollen:

- (1) Normalverteilung: Sie wird zur Beschreibung der Fehlerstruktur metrischer abhängiger Merkmale verwendet und ist die Grundlage von Schätzungen mit der Kleinste-Quadrate-Methode ("ordinary least squares"-Schätzungen in der Regressions-, Varianz- und Kovarianzanalyse).
- (2) Binomialverteilung: Sie wird zur Beschreibung der Auftretenswahrscheinlichkeit eines dichotomen Merkmals in einer Stichprobe verwendet (d.h. des Anteils P_i Untersuchungseinheiten in der i -ten Ausprägung eines dichotomen abhängigen Merkmals). Die Binomialverteilung ist z.B. die Grundlage von Logit-Modellen.
- (3) Poisson-Verteilung: Hat eine nominale abhängige Variable mehr als zwei Ausprägungen, dann sind alternativ Multinomialverteilung oder Poisson-Verteilung zugrunde zu legen. Da sich die Multinomialverteilung unter bestimmten Bedingungen in Produkte von Poisson-Verteilungen überführen läßt (vgl. Arminger 1982, S. 108), verwendet man die weniger komplexe Poisson-Verteilung zur Bestimmung der Auftretenswahrscheinlichkeit einer Stichprobenhäufigkeit in einer Kontingenztafel. Diese Verteilung wird daher im Rahmen log-linearer Modellbildung verwendet.

Die genannten Dichtefunktionen gehören zur Familie der exponentiellen Verteilungen, für die folgende Dichte gilt:

$$F.9 \quad P(Y = y_i) = f(y_i | \theta, \phi) = \exp \left\{ \left[y_i \theta_i - b(\theta_i) \right] / a(\phi) + c(y_i, \phi) \right\} \\ \text{mit } a_i(\phi) = \phi / u_i$$

$f(\cdot)$ ist die Dichte der exponentiellen Familie,

θ_i heißt kanonischer Parameter,

ϕ heißt Dispersionsparameter,

$a(\cdot)$, $B(\cdot)$, $C(\cdot)$ sind geeignet gewählte Funktionen (vgl. Arminger 1982).

Die exponentielle Familie umfaßt neben der stetigen Normalverteilung auch die diskreten Verteilungen, z.B. die Poisson- und Binomialverteilung. Die folgende Tabelle enthält die Charakteristika ausgewählter Verteilungen der exponentiellen Familie:

	<i>Normal</i>	<i>Poisson</i>	<i>Binomial</i>
Wertebereich von y	$(-\infty, \infty)$	$0(1) \infty$	$\frac{0(1)n}{n}$
$a(\phi)$	ϕ	1	$1/n$
$b(\theta)$	$\frac{1}{2}\theta^2$	e^θ	$\ln(1+e^\theta)$
$c(y, \phi)$	$-\frac{1}{2}\left(\frac{y}{\phi} \ln(2\pi\phi)\right)$	$-\ln y$	$\ln\left[\binom{n}{ny}\right]$
$\mu = E(Y)$	θ	e^θ	$e^\theta/(1+e^\theta)$
Varianzfunktion	1	μ	$\mu(1-\mu)$
<ul style="list-style-type: none"> - Normalverteilung: "klassisches lineare Modell" - Poisson-verteilung: "Log-lineares Modell" - Binomialverteilung: "Logit-Modell" 			

In der Theorie verallgemeinerter Modelle geht man somit zunächst von der allgemeinen Dichtefunktion $f(\cdot)$ aus, mit der eine umfangreiche Familie von Exponentialverteilungen gegeben ist, die insbesondere Normal-, Poisson- und Binomialverteilung umfaßt. Diese Funktion hängt von den Parametern Theta (Θ) und Phi (ϕ) ab und enthält die drei Terme $a(\cdot)$, $b(\cdot)$ und $c(\cdot)$, die wiederum Funktionen von Theta, Phi und der abhängigen Variablen Y sind. Diese allgemeine Darstellung gewinnt erst dann praktische Bedeutung, wenn in die drei Terme $a(\cdot)$, $b(\cdot)$ und $c(\cdot)$ der allgemeiner Dichtefunktion bestimmte Umformungen aus der Tabelle eingesetzt werden. Dann ergeben sich u.a. als Spezialfälle die in unserem Zusammenhang wichtigen Verteilungsfunktionen der Normal-, Poisson- und Binomialverteilung. Die verschiedenen Verteilungsannahmen multivariater Analyseverfahren wie Regressionsanalyse mit metrischen Zielvariablen (Normalverteilung), log-lineare Kontingenztabellenanalyse (Poisson-Verteilung) und logistische Regression mit einer dichotomen abhängigen Variablen (Binomialverteilung) sind somit Spezialfälle der verallgemeinerten linearen Modelle. Mit der abhängigen Variablen (Beobachtungswerte einer metrischen Zielvariable, Häufigkeiten einer mehrdimensionalen Kontingenztafel oder Anteilswerte einer dichotomen Zielvariablen) treffen zwar bestimmte Annahmen über die Fehlerverteilung in der Gleichung F.6, der grundlegende Ansatz mit dem linearen Prädiktor

$$\eta = \sum_1^p \beta_j x_j.$$

ändert sich aber nicht. Je nach dem zu analysierenden Datentyp treffen wir eine bestimmte Verteilungsannahme, für die dann eine spezifische Verbindungsfunktion zur Verknüpfung der Modellteile

- linearer Prädiktor und
- Erwartungswert der Beobachtungen $E(y_i) = \mu_i$

gewählt wird. Für die drei Verteilungen sind der folgenden Tabelle die Link-Funk-

tionen zur Verknüpfung der Modellteile aufgeführt:

VERTEILUNG:	LINK-FUNKTION:	BEZEICHNUNG:
Normalverteilung	$\eta = \mu$	IDENTITY
Poissonverteilung	$\eta = \ln \mu$	LOG
Binomialverteilung	$\eta = \ln [\mu / (1 - \mu)]$	LOGIT

(Die korrespondierenden Modellgleichungen und Datenkonstellationen sind auf der folgenden Seite tabelliert).

Durch die Kombination von Verteilungsmodell und Link-Funktion kann das log-lineare Modell im Rahmen des GLM-Ansatzes dargestellt werden. Anstelle der ursprünglichen Beobachtungswerte werden logarithmierte Häufigkeiten erklärt:

$$\ln \mu_i = \eta_i = \beta^T x_i;$$

Die Link-Funktion für das Binomialverteilungsmodell ist unschwer als Logit-Transformation wiederzuerkennen, womit auch der Fall von mehrdimensionalen Kontingenz-tabellenanalysen mit expliziter dichotomer abhängiger Variable seine Darstellung im GLM findet:

$$\log \left(\frac{\pi_i}{1 - \pi_i} \right) = x_i^T \beta$$

$$\pi_i = \frac{e^{x_i^T \beta}}{1 + e^{x_i^T \beta}}$$

Die Häufigkeiten y_i können auch die Werte (0,1) annehmen, so daß nicht nur der Fall aggregierter Daten sondern auch Individualdaten im Rahmen einer logistischen Regression behandelt werden können. Jede "Subpopulation" besteht in diesem Fall aus nur einer Untersuchungseinheit, die Gesamthäufigkeit n_i beträgt somit 1. Die abhängige Variable Y ist praktisch eine Dummy-Variable, die die Zugehörigkeit einer Untersuchungseinheit (i) zu einer der beiden Ausprägungen ($Y_i = 1$) oder ($Y_i = 0$) angibt. Es sei noch einmal darauf aufmerksam gemacht, daß die Umkehrung der Link-Funktionen garantiert, daß für beliebige Werte der Linearkombination η_i der vorgeschriebene Wertebereich der ursprünglichen Zielvariablen eingehalten bleibt. Das bedeutet, daß etwa im logistischen Regressionsmodell die aus der inversen Transformation geschätzten Wahrscheinlichkeiten immer im Intervall (0,1) liegen!

VERBINDUNGSFUNKTIONEN UND MODELLGLEICHUNGEN

$y_i \sim N(\mu_i, \sigma^2)$ $g(\mu_i) = \mu_i$ $\theta_i = \eta_i = \mu_i = \sum_{j=1}^p x_{ij} \beta_j$ $\phi = \sigma^2$	<p>Lineares Modell mit Normalverteilung;</p> <p>link = identity;</p> <p>abhängige Variable = Beobachtungswerte einer metrischen Kriteriumsvariable;</p> <p>- Regressionsanalyse, - Varianz- Kovarianzanalyse.</p>
$y_i \sim P\mu_i$ mit $E(y_i) = \mu_i$ $\theta_i = \eta_i = \ln \mu_i$ $\eta_i = \sum_{j=1}^p x_{ij} \beta_j$ $\phi = 1$	<p>Poissonverteilung (= Multinomialvert., wenn d. x_{ij} Dummy-Variablen sind);</p> <p>link = log-linear;</p> <p>abhängige Variable = Häufigkeit einer Kontingenztafel;</p> <p>- log-lineare Wahrscheinlichkeitsmodelle.</p>
$y_i \sim B(\pi_i, m_i)$ mit $\mu_i = m_i \pi_i$ $\eta_i = g(\mu_i) = \ln(\pi_i / (1 - \pi_i))$ $\eta_i = \sum_{j=1}^p x_{ij} \beta_j$ $\phi = 1$	<p>Binomialverteilung;</p> <p>link = logit;</p> <p>abhängige Variable = Häufigkeit der ersten Ausprägung einer dichotomen abhängigen Variablen;</p> <p>- Logit-Modelle.</p>

Wenn das Verteilungsmodell und die Link-Funktion festgelegt sind, kann die Parameterschätzung erfolgen. Als universelles Schätzverfahren wird die Maximum-Likelihood-Methode zugrundegelegt: Es werden Parameterwerte als Schätzer für die unbekannten β 's der Grundgesamtheit verwendet, die die gegebenen Stichprobenbeobachtungen unter den jeweiligen Modellannahmen am wahrscheinlichsten machen. In dem EDV-Programm GLIM wird zur Berechnung der Parameter β_j die "Scoring"-Methode von Fisher verwendet.

Wir wollen den Schätzlogarithmus hier nicht weiter behandeln und verweisen für weitere Details auf Arminger (1982).

Ist die Berechnung eines Modells beendet, erhält man analog zur Regressionsanalyse neben den Parameterwerten eine geschätzte Varianz-Kovarianz-Matrix der Schätzer, die unter bestimmten Bedingungen normal verteilt sind:

$$V = (X'W'X')^{-1} \quad (V = \text{Varianz-Kovarianz-Matrix von } b, W = \text{Diagonalmatrix der Gewichte}).$$

Mit Hilfe der Standardfehler der Schätzungen (= Wurzel aus den Varianzen) lassen sich dann Tests und Konfidenzintervalle für die einzelnen Parameter konstruieren (s. Abschnitt B).

Neben den Schätzwerten für die Parameter interessiert noch die Güte der Anpassung eines Modells an die beobachteten Daten. Das Ziel einer Analyse ist zunächst, die n Beobachtungswerte einer Kriteriumsvariablen (y_i) durch möglichst wenige unabhängige Variablen (x_{ij} , $j = 1, \dots, p$), d.h. mit wenigen Parametern, zu erklären. Ein hinreichend mit den beobachteten Daten verträgliches Modell, das sowohl die Informationsvielfalt der Daten auf wesentliche Parameter reduziert als auch eine hohe Erklärungskraft aufweist, wird zwischen zwei Extremen liegen:

1. Minimalmodell: Die systematische Komponente ist eine Konstante; zur Vorhersage der Beobachtungswerte wird lediglich der Mittelwert der abhängigen Variablen verwendet, d.h. dieses Modell hat nur einen Parameter. Die Fehlerkomponente ist relativ zu anderen Modellen, in denen weitere Parameter hinzutreten maximal, d.h. die Modellanpassung ist sehr gering.
2. Maximalmodell: In diesem Modell wird $\hat{y}_i = y_i$ gesetzt, d.h. die Anzahl der Parameter ist gleich der Anzahl der Beobachtungswerte. Daher ist die Fehlerkomponente gleich Null und die Modellanpassung somit perfekt.

Um die Güte der Anpassung eines Modells (zwischen beiden Extremen) zu überprüfen, betrachtet man die Devianz die eine Maßzahl für die Abweichung zwischen Modell und Daten ist und auf der Theorie des Likelihood-Ratio-Tests beruht. Die Beurteilung erfolgt im Falle diskreter Daten anhand der CHI^2 -Verteilung, bei der Annahme der

Normalverteilung wird die F-Verteilung verwendet.

Für die drei speziellen Verteilungen erhalten wir folgende Formen der Devianz (summiert wird jeweils über den Index $i = 1 \dots n$):

VERTEILUNG:-----	DEVIANZ:-----
Normalverteilung	$\Sigma(y - \hat{\mu})^2$
Poissonverteilung	$2\{\Sigma[y \ln(y/\hat{\mu}) - (y - \hat{\mu})]\}$
Binomialverteilung	$2\{\Sigma[y \ln(y/\hat{\mu}) + (n - y) \ln((n - y)/(n - \hat{\mu}))]\}$

Das Konzept "Devianz" kann aus der allgemeinen Theorie des Likelihood-Ratio-Tests abgeleitet werden. Bei GLM-Modellen wird ein aktuelles Modell mit p Parametern mit dem Maximalmodell (= saturiertes Modell mit N Parametern, d.h. die Anzahl der Parameter ist gleich der Anzahl der Untersuchungseinheiten) verglichen:

$$\underline{\beta} = [\beta_1, \dots, \beta_p]^T \quad \underline{\beta}_{\max} = [\beta_1, \dots, \beta_N]^T$$

In Termen der Likelihood-Funktion (L) ausgedrückt ist ein Modell den beobachteten Daten hinreichend angepaßt, wenn

$$\text{oder} \quad \begin{aligned} L(\underline{\beta}; \mathbf{y}) &\simeq L(\underline{\beta}_{\max}; \mathbf{y}) \\ L(\underline{\beta}; \mathbf{y}) &\ll L(\underline{\beta}_{\max}; \mathbf{y}) \end{aligned}$$

wenn das Modell dem empirischen Befund nicht gerecht wird. Aus dem Modellvergleich (aktuelles vs. saturiertes Modell) folgt die (verallgemeinerte) Likelihood-Ratio-Statistik als Maß für die Güte der Modellanpassung (in der Notation folgen wir Dobson 1983):

$$\text{F.10} \quad \lambda = \frac{L(\underline{\mathbf{b}}_{\max}; \mathbf{y})}{L(\underline{\mathbf{b}}; \mathbf{y})}$$

$$\text{F.11} \quad \log \lambda = l(\underline{\mathbf{b}}_{\max}; \mathbf{y}) - l(\underline{\mathbf{b}}; \mathbf{y})$$

Liefert (F.11) einen hohen Wert, so ist dies ein Hinweis auf eine unzureichende Modellanpassung.

Aus Gleichung F.11 folgt als Ausdruck für die skalierte Devianz

$$\text{F.12} \quad D = 2 \log \lambda = 2[l(\underline{\mathbf{b}}_{\max}; \mathbf{y}) - l(\underline{\mathbf{b}}; \mathbf{y})],$$

die mit $(N-p)$ Freiheitsgraden χ^2 - verteilt ist. Wir akzeptieren ein Modell, wenn

$$D \sim \chi^2_{N-p}$$

Es wird somit getestet, ob die zusätzlichen Parameter $(p+1, \dots, N)$ im saturierten Modell gegenüber dem aktuellen (eingeschränkten) Modell irrelevant sind. Die Testgröße darf daher keine signifikanten Ergebnisse liefern.

Aus dem Anpassungstest (F.12) kann der klassische Signifikanztest für einzelne oder eine Teilmenge von Parametern abgeleitet werden. Ausgangspunkt kann z.B. das Minimal- oder Basismodell sein, in dem nur die Regressionskonstante spezifiziert ist. Mit Hilfe des multiplen Tests kann überprüft werden, ob die Hinzufügung weiterer unabhängiger Merkmale eine signifikante Modellverbesserung ergibt. Ausgangspunkt eines Signifikanztests sind daher zwei Hypothesen:

$$(1) \text{ Nullhypothese: } H_0: \beta = [\beta_1, \dots, \beta_q]^T = \beta_0$$

und die allgemeinere Alternativhypothese

$$(2) \quad H_1: \beta = [\beta_1, \dots, \beta_p]^T = \beta_1 \quad \text{wobei} \quad q < p < N.$$

Wir testen H_0 gegen H_1 , indem wir die Differenzen der logarithmierten Likelihood-Ratio-Statistiken bilden:

$$\begin{aligned} D &= D_0 - D_1 = 2[l(\mathbf{b}_{\max}; \mathbf{y}) - l(\mathbf{b}_0; \mathbf{y})] - 2[l(\mathbf{b}_{\max}; \mathbf{y}) - l(\mathbf{b}_1; \mathbf{y})] \\ &= 2[l(\mathbf{b}_1; \mathbf{y}) - l(\mathbf{b}_0; \mathbf{y})]. \end{aligned}$$

F.13

Die Differenz der Test-Statistiken (D) ist wiederum χ^2 - verteilt mit $(p-q)$ Freiheitsgraden (q = Anzahl der Parameter im restringierten Modell; p = Anzahl der Parameter im erweiterten Modell).

Als Ergebnis eines Signifikanztests erwarten wir signifikante Werte! Ein Vorteil dieser Teststrategie auf der Grundlage einer Quantifizierung der relativen Modellverbesserung gegenüber einem eingeschränkten Modell ist darin zu sehen, daß dieser Test nicht mehr von der Fallzahl abhängig ist. Wir hatten bereits darauf hingewiesen, daß das Ergebnis eines Anpassungstests auf der Basis des absoluten Devianz-Betrages immer von der Fallzahl abhängig ist.

Die Verwendung der Gleichung F.13 führt zu den Spezialfällen

- (1) Multipler Test für insgesamt p Parameter, wobei

H_0 : Basismodell mit einem Parameter (Regressionskonstante),

H_1 : ein um p Parameter erweitertes Modell.

- (2) Partieller Test für einen Parameter aus der Menge $(\beta_1, \dots, \beta_p)$ oder einer Teilmenge, wobei

H_0 : allgemein ein hierarchisches Modell mit einer Teilmenge von Parametern aus $(\beta_1, \dots, \beta_p)$

H_1 : ein Modell mit p Parametern.

Wir wollen im folgenden noch kurz auf die Devianz-Berechnung für spezielle Submodelle des GLM-Ansatzes eingehen:

- (1) NORMALVERTEILUNG:

Unter der Annahme, daß die Beobachtungswerte Y_1, \dots, Y_N der abhängigen Variablen voneinander unabhängig und normal verteilt sind lautete die Likelihood-Funktion

$$F.14 \quad l(\beta; y) = -\frac{1}{2\sigma^2} \sum_{i=1}^N [y_i - E(Y_i)]^2 - \frac{1}{2} N \log(2\pi\sigma^2)$$

Für das Maximalmodell erhalten wir den Ausdruck

$$F.15 \quad E(Y_i) = \mu_i, i = 1, \dots, N \quad \beta_{\max} = [\mu_1, \dots, \mu_N]^T \quad \hat{\mu}_i = y_i$$

$$l(\mathbf{b}_{\max}; y) = -\frac{1}{2} N \log(2\pi\sigma^2)$$

Für das Basismodell erhalten wir den Ausdruck

$$F.16 \quad E(Y_i) = \mu \text{ für alle } i \quad \beta = [\mu]^T \quad \hat{\mu} = \bar{y}.$$

$$l(\mathbf{b}; y) = -\frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - \bar{y})^2 - \frac{1}{2} N \log(2\pi\sigma^2)$$

Daher erhalten wir nach Gleichung F.12 für die skalierte Devianz des Basismodells

$$F.17 \quad D_0 = \frac{1}{\sigma^2} \sum_{i=1}^N (y_i - \bar{y})^2.$$

$\beta = (\beta_1, \dots, \beta_p)^T$: (aktuelle oder laufende) Modell $E(\underline{Y}) = \underline{X}\beta$ mit p Parametern
 $\underline{\beta} = (\beta_1, \dots, \beta_p)^T$

$$F.18 \quad D_1 = 2[l(\mathbf{b}_{\max}; y) - l(\mathbf{b}; y)] \\ = \frac{1}{\sigma^2} (y - \mathbf{X}\mathbf{b})^T (y - \mathbf{X}\mathbf{b})$$

Da der Wert von σ^2 in der Regel unbekannt ist, sind die Gleichungen F.17 und F.18 nicht bestimmbar. Wenn man jedoch anstelle der absoluten skalierten

Devianzen die Verhältnisse der skalierten Devianzen betrachtet, kann man den unbekannten Parameter σ^2 kürzen. Den Anteil erklärter Varianz in der klassischen Regressionsanalyse erhalten wir durch folgenden Modellvergleich:

$$F.19 \quad R^2 = (D_0 - D_1) / D_0$$

Durch den Bruch wird σ^2 weggekürzt!

Da der Quotient zweier χ^2 -verteilter Größen, die jeweils durch ihre Freiheitsgrade geteilt wurden, einer F-Verteilung genügen, erhalten wir aus dem Modellvergleich D_0 versus D_1 die Prüfgröße oder Regressionsanalyse:

$$F.20 \quad f = \frac{(D_0 - D_1) / D_1}{(p - q) / N - p} \sim F_{p-q, N-p}$$

(2) BIX OMI AL VERTEILUNG:

Unter der Annahme der Binomialverteilung erhalten wir das logistische Regressionsmodell mit einer dichotomen abhängigen Variablen mit der Gleichung.

$$\text{logit } \pi_i = \log\left(\frac{\pi_i}{1 - \pi_i}\right) = \mathbf{x}_i^T \boldsymbol{\beta} \quad , i = 1, \dots, N ,$$

wobei (X) = Anzahl der Untersuchungseinheiten in der Stichprobe, wenn das Modell auf der Basis von Individualdaten geschätzt wird und

(N) = Anzahl der Subpopulationen, wenn das Modell auf der Basis aggregierter Daten geschätzt wird.

Um die Güte der Modellanpassung quantifizieren zu können, gehen wir von folgendem Ausdruck aus:

$$F.21 \quad D = 2[l(\hat{\pi}_{\max}; \mathbf{r}) - l(\hat{\pi}_0; \mathbf{r})]$$

$$\mathbf{r} = [r_1, \dots, r_N]^T, \hat{\pi}_0 = [\hat{\pi}_1, \dots, \hat{\pi}_N]^T$$

gleich dem Vektor der ML-Schätzungen des Maximalmodells (max) bzw. des aktuellen - reduzierten - Modells (o).

Der logarithmierte Likelihood-Wert des Maximalmodells lautet:

$$F.22 \quad l(\hat{\pi}_{\max}; \mathbf{r}) = \sum_{i=1}^N \left[r_i \log\left(\frac{r_i}{n_i}\right) + (n_i - r_i) \log\left(\frac{n_i - r_i}{n_i}\right) + \log(n_i) \right]$$

mit (r_i / n_i) = beobachteten Anteil in der Ausprägung ($Y_i = 1$) für die i-te Untersuchungseinheit bzw. Subpopulation bei aggregierten Daten.

In dem saturierten Modell werden die Beobachtungswerte perfekt reproduziert,

d.h. die geschätzten Werte sind gleich den beobachteten Werten.

Nach Gleichung F.21 folgt nun die Devianz des aktuellen Modells aus

$$F.23 \quad D = 2 \sum_{i=1}^N \left[r_i \log \left(\frac{r_i}{n_i \hat{\pi}_i} \right) + (n_i - r_i) \log \left(\frac{n_i - r_i}{n_i - n_i \hat{\pi}_i} \right) \right].$$

Somit hat der Ausdruck für D die Form

$$F.24 \quad D = 2 \sum o \log \frac{o}{e}$$

die äquivalent mit der bereits im ersten Abschnitt erläuterten Gleichung für die CHI_L^2 -Statistik ist.

CHI_L^2 folgt mit $(n-p)$ einer \hat{X}^2 -Verteilung; "o" symbolisiert die beobachteten Häufigkeiten r_i bzw. $n_i - r_i$ und "e" symbolisiert die korrespondierenden Schätzwerte unter der jeweiligen Parameterspezifikation im aktuellen Modell.

(3) POISSON-VERTEILUNG.

Unter der Annahme der Poisson-Verteilung liegt eine log-lineare Kontingenztabellenanalyse vor. Zur Berechnung der Devianz vergleichen wir wiederum das saturierte Modell, das die empirischen Zellenhäufigkeiten perfekt reproduziert, mit dem laufenden oder aktuellen Modell. Die Werte der logarithmierten Likelihood-Funktionen entnehmen wir in diesem speziellen Fall den Gleichungen

$$l(\mathbf{b}_{\max}; \mathbf{y}) = \text{constant} + \sum_i y_i \log y_i.$$

$$l(\mathbf{b}; \mathbf{y}) = \text{constant} + \sum_i y_i \log e_i.$$

Daher folgt die logarithmierte Likelihood-Ratio-Statistik aus

$$F.25 \quad D = 2[l(\mathbf{b}_{\max}; \mathbf{y}) - l(\mathbf{b}; \mathbf{y})] = 2 \sum_{i=1}^N y_i \log \frac{y_i}{e_i}$$

die analog zum logistischen Modellansatz die Form

$$F.26 \quad D = 2 \sum o \log \frac{o}{e}$$

annimmt. Die Devianz ist somit auch in diesem speziellen Submodell des GLM das bekannte Testkriterium des log-linearen Modells:

$$D - \text{CHI}_L^2 = G'$$

Mit der Gegenüberstellung der Devianzen ausgewählter spezieller Verteilungen im Rahmen des verallgemeinerten linearen Modells sollte deutlich werden, daß das Konzept "Devianz" jeweils mit bekannten Maßzahlen zur Beurteilung der Güte der Modellanpassung an den empirischen Befund in speziellen multivariaten Analyseverfahren identisch ist. Die Submodelle des GLM legen lediglich spezielle Testkriterien fest. Grundsätzlich bleibt die Interpretation der Devianz als Abweichung der empirischen von den modellimplizierten (d.h. prognostizierten) Werten als allgemeiner Ansatz erhalten.

Aus der Devianz lassen sich die bereits diskutierten partiellen und multiplen Bestimmtheitsmaße für das jeweilige Submodell des GLM ableiten, wobei die Interpretation unabhängig von der speziellen Verteilungsannahme erhalten bleibt:

$$\text{F.27} \quad R^2 = (D_0 - D_1) / D_0,$$

$$\text{F.28} \quad sr^2 = (D_{M_1} - D_{M_2}) / D_0,$$

$$\text{F.29} \quad \text{PRD} = (D_{M_1} - D_{M_2}) / D_{M_2},$$

wobei (M_1) jeweils ein reduziertes Modell ist, dessen Parameter eine Teilmenge des erweiterten Modells (M_2) ist.

Wir hatten bereits festgestellt, daß die systematische Komponente μ der Zielvariablen Y Gegenstand des eigentlichen Erklärungsmodells ist. Mit Hilfe der Link-Funktion (g) wird sie mit dem linearen Prädiktor η verbunden, der die unabhängigen Merkmale zusammenfaßt. Wir wollen den modelltheoretischen Teil des GLM-Ansatzes im folgenden mit der Frage abschließen, wie dieser lineare Prädiktor im Fall nominaler Merkmale bestimmt wird (Modellparametrisierung, s. a. Abschnitt A und B).

Sind die unabhängigen Variablen nominal/ordinal skaliert, stellt sich etwa im Rahmen einer Varianz-, Logit- oder log-linearen Analyse die Frage nach einer geeigneten Transformation, so daß die Merkmale in die Design-Matrix \underline{x} des verallgemeinerten Modells aufgenommen werden können. Dazu müssen die einzelnen Merkmalskategorien einer diskreten Variablen in 'technische' Prädiktorvariablen übersetzt werden, die nur zwei Ausprägungen haben: Zugehörigkeit einer Untersuchungseinheit (numerisch dargestellt durch den Wert "1") und Nichtzugehörigkeit (numerisch dargestellt durch den Wert "0") zu einer Merkmalsausprägung. Da sich die Ausprägungen gegenseitig ausschließen, kann pro Untersuchungseinheit (= Zeile der Designmatrix) nur jeweils eine der Kodiervariablen (= Spalte der Designmatrix) den Wert "1" haben. Diese 0/1-kodierten Variablen haben wir bereits unter dem Begriff "Dummy-Variablen" eingeführt. Sie werden von dem EDV-Programm GLIM immer dann erzeugt, wenn unabhängige Merkmale als diskret (in der GLIM-Syntax als FACTOR) deklariert werden.

Betrachten wir als Beispiel folgende zweidimensionale Häufigkeitstabelle (AxB) mit den Häufigkeiten $n_{ij} = y_{ij}$ als abhängige Größe:

	B ₁	B ₂	B ₃
A ₁	y ₁₁	y ₁₂	y ₁₃
A ₂	y ₂₁	y ₂₂	y ₂₃

$$Y = \begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{bmatrix}$$

mit A_i , i = 1,2 und B_j , j = 1,2,3

Die Zellhäufigkeiten können mit der Poisson-Verteilung beschrieben werden. Das saturierte log-lineare Modell läßt sich dann im Rahmen des GLM-Ansatzes wie folgt spezifizieren:

$$y_{ij} = \mu + e_{ij} \quad \text{wobei } e_{ij} \text{ Poisson-verteilt ist, } i = 1,2 \text{ und } j = 1,2,3,$$

Verbindungs-Funktion: log-linear,

$$y_{ij} = \exp(\eta_{ij}) + e_{ij},$$

$$\eta_{ij} = \beta_0 + \beta_i^A + \beta_j^B + \beta_{ij}^{AB} \quad (\text{Gleichung des saturierten Modells}),$$

mit β_0 = Grundeffekt; β_i^A und β_j^B = Haupteffekte; β_{ij}^{AB} = Interaktionseffekt.

Unter Verwendung von Dummy-Variablen nimmt die Designmatrix \underline{x} folgende Formen an:

AUSPRÄGUNGS= KOMBINATION	D ₁ ^A	D ₂ ^A	D ₁ ^B	D ₂ ^B	D ₃ ^B	D ₁₁ ^{AB}	D ₁₂ ^{AB}	D ₁₃ ^{AB}	D ₂₁ ^{AB}	D ₂₂ ^{AB}	D ₂₃ ^{AB}
A ₁ B ₁ →	1	0	1	0	0	1	0	0	0	0	0
1 2 →	1	0	0	1	0	0	1	0	0	0	0
1 3 →	1	0	0	0	1	0	0	1	0	0	0
2 1 →	0	1	1	0	0	0	0	0	1	0	0
2 2 →	0	1	0	1	0	0	0	0	0	1	0
2 3 →	0	1	0	0	1	0	0	0	0	0	1

Eine genaue Betrachtung dieses Ansatzes zeigt, daß die Forderung nach linearer Unabhängigkeit der Spalten der Designmatrix voneinander hier nicht erfüllt ist. So ist

etwa die Summe der ersten zwei Spalten gleich der Summe der dritten bis fünften Spalte. Dieses Problem wird in dem GLM-Ansatz durch die lineare Restriktion "cornered effects" (als Reparametrisierungsbedingung) gelöst. Pro Variable wird jeweils der erste Effekt gleich Null gesetzt, d.h. für unser Beispiel gilt:

$$\beta_1^A = \beta_1^B = \beta_{11}^{AB} = \beta_{12}^{AB} = \beta_{13}^{AB} = \beta_{21}^{AB} = 0$$

Das saturierte log-lineare Modell für die zweidimensionale Kontingenztafel kann dann im Rahmen des GLM-Ansatzes wie folgt beschrieben werden:

$$\begin{array}{l}
 E y_{1j} = \mu_{1j}, \eta_{1j} = \ln \mu_{1j} = \beta_0 + \beta_1^A + \beta_2^B + \beta_{1j}^{AB} \\
 \text{mit } \beta_1^A = \beta_1^B = \beta_{11}^{AB} = \beta_{12}^{AB} = \beta_{13}^{AB} = \beta_{21}^{AB} = 0
 \end{array}$$

	β_0	β_1^A	β_2^B	β_3^B	β_{22}^{AB}	β_{23}^{AB}	
η_{11}	1	0	0	0	0	0	β_0
η_{12}	1	0	1	0	0	0	β_2^A
η_{13}	1	0	0	1	0	0	β_2^B
η_{21}	1	1	0	0	0	0	β_3^B
η_{22}	1	1	1	0	1	0	β_{22}^{AB}
η_{23}	1	1	0	1	0	1	β_{23}^{AB}

Da die Dummy-Variablen für die Interaktionseffekte als Produkte der Dummy-Variablen der Haupteffekt gebildet werden, sind auch die Interaktionseffekte, an denen eine erste Ausprägung beteiligt ist, gleich Null.

Die Besonderheit der zweiten Reparametrisierungsbedingung ergibt sich aus dem konstanten Glied (β_0), das nun nicht mehr als Mittelwert interpretiert werden kann, sondern eine nicht weiter zerlegbare Kombination der jeweils ersten Ausprägungen aller nominalskalierten Merkmale darstellt, einschließlich der multiplikativ gebildeten Interaktionen, in denen die Ausprägung (1) auftritt. In dem konstanten Glied sind alle ursprünglich getrennt bestehenden Kodiervariablen zusammengefaßt, die jeweils die erste Ausprägung einer Variablen repräsentieren (einschließlich der Dummy-Variablen für die Interaktionen A, B , A, B , A, B , A, B). Demnach sind die spezifischen Einflußgewichte der gleich Null gesetzten Effekte nicht mehr voneinander zu unterscheiden, da sie gemeinsam in dem konstanten Glied repräsentiert sind. Die Einflußgewichte für die verbleibenden Ausprägungen und deren Interaktionen werden relativ zum Grundeffect (β_0) bestimmt. Sie sind daher auch als Abweichungen von dem konstanten Ausgangswert zu interpretieren. Die Bezeichnung "cornered effects" (auf

eine Kategorie bezogene Effekte) charakterisieren die Interpretation der Effekte als Abweichung von der ersten (Basis-) Kategorie aller Merkmale.

Der Haupteffekt von A sagt bei dieser Reparametrisierung aus, um wieviel die Beobachtungen der (transformierten) abhängigen Variablen über der Vergleichskategorie liegen, die ausschließlich den Wert A_2 aufweisen. Analog sind die Haupteffekte der Kategorien B_2 und B_3 zu interpretieren. Die Interaktionseffekte drücken schließlich aus, um wieviel die Beobachtungen über der Basiskategorie (A_1B_1) liegen, wenn sowohl der Wert A_i ($i \neq 1$) als auch der Wert B_j ($j \neq 1$) beobachtet wird, vorausgesetzt diese Abweichungen wird nicht schon durch die Haupteffekte erfaßt.

Die etwas umständliche Interpretation wird anschaulicher, wenn wir die Werte der abhängigen Variablen (n^{\wedge}) aufgrund der Informationen der unabhängigen Prädikatoren A und B vorhersagen. Dazu benötigen wir die Umkehrung der Link-Funktion für das log-lineare Modell:

$$y_{ij} = g^{-1}(\eta_{ij}) + e_{ij} \quad ; \quad y_{ij} = e^{\eta_{ij}} + e_{ij}$$

bzw. $y_{ij} = \exp(\eta_{ij}) + e_{ij}$, wobei $e_{ij} = 0$ im saturierten Modell.

Die folgende Tabelle enthält die Verteilung der Parameter des saturierten log-linearen Modells, die für eine zellenspezifische Vorhersage der einzelnen Häufigkeiten unter Berücksichtigung der Reparametrisierungsbedingungen relevant sind, auf die entsprechenden Zellen der (AxB)-Kontingenztafel. Eine Modellvorhersage setzt sich nach Transformation von η_{ij} aus der Summe der angegebenen Parameter zusammen:

$y_{ij} = \exp(\beta_0 + \beta_1^A + \beta_j^B + \beta_{ij}^{AB})$			
	B_1	B_2	B_3
A_1	$y_{11} = \exp(\beta_0)$	$y_{12} = \exp(\beta_0 + \beta_2^B)$	$y_{13} = \exp(\beta_0 + \beta_3^B)$
A_2	$y_{21} = \exp(\beta_0 + \beta_1^A)$	$y_{22} = \exp(\beta_0 + \beta_2^A + \beta_2^B + \beta_{22}^{AB})$	$y_{23} = \exp(\beta_0 + \beta_2^A + \beta_3^B + \beta_{23}^{AB})$

Das Schema demonstriert, daß das saturierte Modell bei geeigneter Wahl der Parameterwerte in der Lage ist, jede beobachtete Häufigkeit einer Ausprägungskombination (A,B) als Modellvorhersage zu liefern.

Obwohl sich die Interpretation der Parameter je nach Wahl der Reparametrisierungsbedingung unterscheidet, liegen die Parameterdifferenzen zwischen jeweils zwei Kategorien unabhängig von der Reparametrisierung eindeutig fest (vgl. Arminger 1982,

S. 55). Für den Unterschied zwischen zwei beliebigen Kategorien j und k eines Merkmals mit m Ausprägungen gilt die Beziehung

$$F.30 \quad \beta_j^c - \beta_k^c = \beta_j^z - \beta_k^z \quad , \quad j, k \text{ aus } 1, \dots, m.$$

(c = cornered effects, z = centered effects)

Wir wollen den theoretischen Teil mit einer Diskussion der Vor- und Nachteile abschließen.

- Die Interpretation der Effekte als Abweichungen vom Gesamtdurchschnitt ist bei der (1/-1)-Kodierung im Vergleich zu 'centered effects' sehr anschaulich. Die leichte Interpretierbarkeit geht aber bei ungleichen Zellenbesetzungen (etwa in der Varianzanalyse) oder unbesetzten Zellen in einer Kontingenztafel verloren. Auf eine Kategorie bezogene Effekte sind dagegen von der gegebenen Datenkonstellation unabhängig.
- Der Übergang von log-linearen Modellen zu Logit-Modellen mit einer polytomen abhängigen Variablen ist im Hinblick auf die Bestimmung der Logit-Parameter - wie noch zu zeigen sein wird - besonders einfach.
- Unbesetzte Zellen können im Rahmen des GLM-Ansatzes wesentlich flexibler behandelt werden (s.d. Diskussion in Aston/Wilson 1984, S. 71ff.). Die Anwendung alternativer Verfahren zur multivariaten Analyse nominaler Merkmale (z.B. ECTA, BMDP4F, NONMET) erfordern in diesem Fall eine Vorgabe von künstlichen Werten für unbesetzte Zellen, um die Haupt- und Interaktionseffekte schätzen zu können. Diese Vorgehensweise führt zu einer falschen Anzahl von Freiheitsgraden (s. Arminger 1983, S. 49) und gestattet keine Trennung zwischen
 1. 'sampling zeros' (die Stichprobe ist nicht groß genug, so daß nicht alle Merkmalskombinationen besetzt sind) und
 2. strukturellen Nullen (bestimmte Merkmalskombinationen sind inhaltlich nicht sinnvoll oder garnicht zugelassen),

da in beiden Fällen künstliche Daten für die unbesetzten Zellen eingesetzt werden. Das Programm GLIM ermöglicht dem Anwender hingegen eine differenzierte Behandlung von Ausprägungskombinationen, für die Beobachtungen fehlen:

1. 'sampling zeros': Man geht von einer vollständigen Datenmatrix aus und bezieht die unbesetzten Zellen mit in die Analyse ein. In diesem Spezialfall werden einige Parameter fehlerhaft berechnet. Dies läßt sich leicht anhand der Standardfehler erkennen, deren Werte relativ zu den Parameterschätzungen sehr groß werden (vgl. Wilson/Aston 1984, S. 72). Bei der Interpretation des empirischen Befundes werden diese Parameter ausge-

geschlossen.

2. Strukturelle Nullen: Die unbesetzten Ausprägungskombinationen (= Zeilen der Designmatrix) werden von der Analyse ausgeschlossen. Damit hat die Designmatrix keinen vollen Spaltenrang mehr und die Parameter können nicht berechnet werden. Bestimmte, nicht berechenbare Parameter werden daher von dem Programm gleich Null gesetzt und als "aliased" ausgewiesen.

Wir wollen die Behandlung fehlender Werte anhand der obigen (2x3)-Kontingenztafel illustrieren (vgl. Arminger 1982, S. 62):

$$\begin{bmatrix} n_{11} \\ n_{12} \\ n_{13} \\ n_{21} \\ n_{22} \\ n_{23} \end{bmatrix} = \begin{bmatrix} \beta_0 & \beta_2^A & \beta_2^B & \beta_3^B & \beta_{22}^{AB} & \beta_{23}^{AB} \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_{A2} \\ \beta_{B2} \\ \beta_{B3} \\ \beta_{A2B2} \\ \beta_{A2B3} \end{bmatrix}$$

Wenn nun die Zellen (A = 1 und B = 2) unbesetzt ist, wird die entsprechende Zeile der Designmatrix gestrichen. Die Matrix \underline{x} ohne die zweite Zeile lautete dann:

$$\underline{x} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

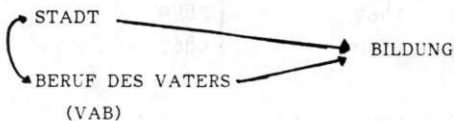
Die fünfte Spalte (die mit β_{22}^{AB} multipliziert wird) ist dann mit der dritten Spalte identisch. Somit ist der Rang der Matrix \underline{x} kleiner als die Anzahl der zu schätzenden Parameter. Anders ausgedrückt: die dritte und fünfte Spalte bilden einer perfekte Linearkombination, sie sind also nicht unabhängig voneinander. Als Folge der unbesetzten Zelle kann der Parameter β_{22}^{AB} nicht mehr geschätzt werden, wir setzen diesen Wert daher gleich Null. In dem EDV-Programm GLIM würde die entsprechende Spalte der Designmatrix gestrichen werden und der korrespondierende Parameter als "aliased" ausgewiesen.

Man beachte außerdem, daß nach Wegfall der dritten Zeile der Designmatrix das Design nicht mehr orthogonal ist. Eine Effektkodierung (-1/+1) der Stellvertreter-

variablen ergäbe keine zentrierten Effekte mit der Interpretation "Abweichung vom Gesamtdurchschnitt"! Bei unvollständigen Designs oder Daten mit Stichprobennullen ist daher die Dummy-Kodierung vorteilhafter.

Anknüpfend an unsere theoretischen Überlegungen wollen wir im folgenden anhand eines Beispiels zeigen, wie die Logit-Analyse und log-lineare Analyse als Spezialfälle verallgemeinerter linearer Modelle mit Hilfe des EDV-Programms GLIM umgesetzt wird. Dabei zeigen wir neben der Modellspezifikation die Interpretation der einzelnen Ergebnisse im Rahmen der GLIM-Analyse.

Wir gehen von folgendem Abhängigkeitsmodell aus:



Uns interessiert die Frage, wie sich das Auftreten bestimmter Bildungsränge durch die unabhängigen Variablen STADT und VAB prognostizieren läßt. Da die Kategorien unserer abhängigen Variablen geordnet sind, ist es sinnvoll, die in Abschnitt (B) diskutierten "continuation ratios" zu verwenden. Wir betrachten daher

1. die Wahrscheinlichkeit, in die erste Kategorie B(1) zu kommen im Verhältnis zur Wahrscheinlichkeit, in eine der beiden Restkategorien zu fallen, und
2. die Wahrscheinlichkeit, in die zweite Kategorie B(2) zu kommen im Verhältnis zur Wahrscheinlichkeit, in die dritte Kategorie zu fallen,

als abhängige Variablen. Die genauen Daten über die Variablenausprägungen und ihre Häufigkeiten sind in den Tabellen auf der nächsten Seite dargestellt.

Unser Ziel ist es, die Wahrscheinlichkeit eines bestimmten Bildungsranges im Vergleich zu den jeweils nächst höheren Bildungsrängen durch einen linearen Prädiktor zu erklären, der sich aus den Ausprägungen der beiden Merkmale VAB und STADT und deren Interaktionen zusammensetzt. Für eine dichotome abhängige Variable lautet die Modellspezifikation:

Verbindungsfunktion = Logit, Zielvariable = binomialverteilt, d.h. wir betrachten zwei Logit-Modelle mit den abhängigen Variablen

$$1. \eta_i(1) = \ln(R_{i1} / N_{ij1} - R_{i1}) ,$$

$$2. \eta_i(2) = \ln(R_{i2} / N_{ij2} - R_{i2}) ,$$

für jede Merkmalskombination VAB/STADT ($i = 1, \dots, 6$),

R_{i1} = Zahl der Volksschüler in der Merkmalskombination VAB/STADT,

R_{i2} = Zahl der Schüler mit einem Bildungsrang "Mittelschule - Tertia" in der Merkmalskombination VAB/STADT.

BASISTABELLEN DER LOGIT-MODELLE MIT "CONTINUATION RATIOS"

TABELLE 1: STADT	VAB	BILDUNGSRANG = VOLKSSCHULE	N_{1j1}
1	1	10	196
1	2	394	1536
1	3	1580	2050
2	1	24	148
2	2	1969	2849
2	3	6578	6867

TABELLE 2: STADT	VAB	BILDUNGSRANG =MITTELSCH.-TERTIA	N_{1j2}
1	1	118	186
1	2	891	1142
1	3	449	470
2	1	64	124
2	2	633	880
2	3	270	289

N_{1j1} : Zahl der Schulabgänger insgesamt (in der Ausprägungskombination S_1V_j);

N_{1j2} : Zahl der Schulabgänger mit einem Bildungsrang zwischen Mittelschule bis Prima (in der Ausprägungskombination S_1V_j).

Ausprägungen der Variablen

VAB: (1) Besitz- und Bildungsbürger
(2) alter und neuer Mittelstand
(3) Arbeiter und untere Beamte

STADT: (1) Minden
(2) Duisburg

Die Ergebnisse für die beiden GL-Modelle sind in der Tabelle auf S.130 dargestellt. Zunächst wird ein passendes Modell für die Daten gesucht. Dazu betrachten wir die erklärten Devianzen für unterschiedliche Modelle. Für die zwei Logit-Modelle gilt, daß die Haupteffekte jeweils für 99% der gesamten Devianz aufkommen. Die Vernachlässigung jedes einzelnen Haupteffekts hat eine bedeutsame Zunahme der Devianz zur Folge. Es ergeben sich aber unterschiedliche Ergebnisse im Hinblick auf die relativen Erklärungsanteile. Während die Vernachlässigung der Variablen STADT im ersten Logit-Modell 30% weniger erklärte Devianz zur Folge hat (= relativer Erklärungsanteil dieser Variablen), beobachten wir im zweiten Logit-Modell lediglich eine Reduktion der erklärten Devianz von 6%. Für das (Logit-) Verhältnis "Mittelschule - Tertia" versus "Sekunda - Prima" ist die Variable STADT offensichtlich ein relativ unbedeutsamer Prädiktor.

Betrachten wir noch die einzelnen Parameterwerte für die getrennt geschätzten Haupteffektmodelle. Hier zeigen sich erhebliche Unterschiede in der Bedeutsamkeit einzelner Kategorien für die Vorhersage der Zielvariablen. Die Wahrscheinlichkeit, einen Volksschulabschluß zu erreichen (im Vergleich zu den zusammengefaßten höheren Bildungsrängen), wird durch die Ausprägung STADT = "Duisburg" erheblich erhöht. Der Effekt dieser Ausprägung auf die Wahrscheinlichkeit eines Bildungsranges "Mittelschule - Tertia" (im Vergleich zu "Sekunda - Prima") ist dagegen unbedeutend, d.h. die Variable STADT diskriminiert nur zwischen einem Volksschulabschluß und allen anderen Bildungsrängen.

In beiden Modellen erweist sich die dritte Kategorie der Variablen VAB als der bedeutsamste Prädiktor, d.h. die Wahrscheinlichkeit für einen niedrigen Bildungsrang wird im Vergleich zu den nächst höheren Bildungsrängen für Arbeiter und untere Beamte deutlich erhöht.

Interessant ist noch ein Größenvergleich des Effektes der Kategorie VAB = "alter und neuer Mittelstand". Die Wahrscheinlichkeit eines Volksschulabschlusses (im Vergleich zu allen höheren Bildungsrängen) wird durch die Ausprägung VAB(2) beträchtlich erhöht. Dieser Effekt ist dagegen in dem zweiten Logit-Modell erheblich niedriger, d.h. die Ausprägung VAB(2) diskriminiert nicht zwischen den Bildungsrängen "Mittelschule-Tertia" und "Sekunda - Prima".

Bisher haben wir lediglich die Vorzeichen einzelner Effekte interpretiert und deren Größenordnung untereinander verglichen, denn es gibt keine so anschauliche Interpretation der numerischen Werte einzelner Koeffizienten wie im allgemeinen linearen Modell mit einer metrischen Zielvariablen. Wir können jedoch die Modellgleichung

$$\log \left(\frac{\hat{p}_i}{1 - \hat{p}_i} \right) = \beta_0 + \sum_{j=1}^r \beta_j z_{ij}$$

(z_{ij} = Wert der j-ten Dummy-Variablen in dem spezifizierten Modell in der i-ten Zeile der Designmatrix).

ERGEBNISSE DER LOGIT-ANALYSEN MIT "CONTINUATION RATIOS"

MODELLE ZUR ERKLÄRUNG DER DATEN IN TABELLE 1:

MODELL	DEVIANZ	FG	ERKLÄRTE DEVIANZ
STADT + VAB	2,331	2	0,99
STADT	2811,0	4	0,38
VAB	1380,0	3	0,69
BASISMODELL (GM)	4516,0	5	0,0

SCHÄTZUNGEN DER PARAMETER DES HAUPTEFFEKTMODELLS (TABELLE 1):

ESTIMATE	S.E.	PARAMETER
- 3,360	0,19	%GM
1,879	0,05	STADT(2)
2,289	0,19	VAB(2)
4,587	0,19	VAB(3)

MODELLE ZUR ERKLÄRUNG DER DATEN IN TABELLE 2:

MODELL	DEVIANZ	FG	ERKLÄRTE DEVIANZ
STADT + VAB	0,419	2	0,99
STADT	228,0	4	0,07
VAB	15,71	3	0,93
BASISMODELL (GM)	246,1	5	0,0

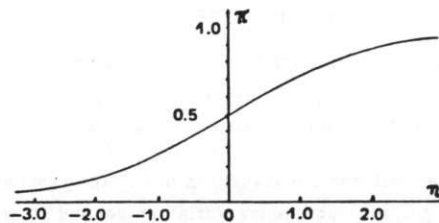
SCHÄTZUNGEN DER PARAMETER DES HAUPTEFFEKTMODELLS (TABELLE 2):

ESTIMATE	S.E.	PARAMETER
0,49	0,12	%GM
- 0,35	0,09	STADT(2)
0,78	0,13	VAB(2)
2,54	0,20	VAB(3)

für die Vorhersage einzelner Auftretenswahrscheinlichkeiten P_i (bzw. $\pi_i = 1, \dots, 6$) verwenden. Die prognostizierten Wahrscheinlichkeiten erhalten wir aus der Umkehrtransformation

$$F.31 \quad \pi_i = \exp(\hat{\eta}_i) / (1 + \exp(\hat{\eta}_i))$$

Der Zusammenhang zwischen η und π wird durch folgendes Schaubild illustriert (Quelle: Arminger 1982, S. 51):



(Graph der Funktion $\pi_i = \exp(\eta_i) / (1 + \exp(\eta_i))$).

Die folgende Tabelle zeigt die Ergebnisse einzelner Prognosen für unterschiedliche Ausprägungskombinationen der unabhängigen Variablen (Basis: Haupteffektmodell mit der abhängigen Variablen Y = Volksschulabschluß versus höhere Bildungsstufen):

AUSPRÄGUNGSKOMBINATION $S_i V_j$	$\eta_i = \beta_0 + \beta_j^V + \beta_i^S$	PROGNOSE π_{iA}
VAB = 1; STADT = 1 (Vergleichsgruppe)	- 3,36	0,033
VAB = 1; STADT = 2	- 3,36 + 1,88 = - 1,48	0,187
VAB = 2; STADT = 1	- 3,36 + 2,29 = - 1,07	0,255
VAB = 2; STADT = 2	- 3,36 + 2,29 + 1,88 = 0,81	0,692

Die Regressionskonstante (GM = -3,36) gibt - nach Transformation - die Wahrscheinlichkeit oder den Anteilswert an, daß der Sohn eines Besitz- oder Bildungsbürgers in Minden einen Volksschulabschluß hat. Der Anteil beträgt in dieser Vergleichsgruppe

(Ausprägungskombination S_1V_1) 3% Im Vergleich zu $S(1)$ - Minden - wird die Wahrscheinlichkeit für einen Volksschulabschluß erhöht, wenn der Sohn in Duisburg zur Schule ging. Der prognostizierte Anteil beträgt in dieser Ausprägungskombination (V_1, S_1) 18,7%. Betrachtet man hingegen die Herkunftsschicht "alter und neuer Mittelstand" ($VAB = 2$) im Vergleich zu $VAB = 1$ und die Stadt Duisburg im Vergleich zu Minden (S_1), dann erhöht sich der Anteilswert beträchtlich; er beträgt nun 69,2%. Als letzte Beispiel berechnen wir noch die Wahrscheinlichkeit eines Volksschulabschlusses, wenn der Vater Arbeiter oder unterer Beamter ($VAB = 3$) in Duisburg ($STADT = 2$) war:

$$\begin{aligned}\eta &= GM + \beta_3^V + \beta_2^S \\ \eta &= -3,36 + 4,587 + 1,879 = 3,106 \\ \pi &= \exp(3,106) / (1 + \exp(3,106)) = 22,33/23,33 = 0,957\end{aligned}$$

Die Parameter β_3^V und β_2^S quantifizieren jeweils die Abweichungen von der Vergleichsgruppe (VAB_1 , $STADT_1$), d.h. sie erhöhen in diesem Beispiel jeweils die Wahrscheinlichkeit relativ zur Vergleichsgruppe um 4,5 bzw. 1,8 Punkte.

Wir wollen uns im Folgenden mit der Beantwortung der Frage beschäftigen, wie im Rahmen des GLM-Ansatzes polytome abhängige Variablen behandelt werden. Wie muß ein Modell für die abhängige Variable BILDUNG spezifiziert werden, wenn wir nicht von "continuation ratios" ausgehen (Zerlegung der abhängigen Variablen in zwei dichotome Merkmale), sondern die ursprünglich drei Ausprägungen beibehalten wollen? In Abschnitt (B) haben wir bereits gezeigt, wie sich Logit-Modelle auf polytome abhängige Variable verallgemeinern lassen. Da ein Logit-Modell aus einem speziellen log-linearen Modell abgeleitet werden kann, können wir den ursprünglich symmetrischen log-linearen Ansatz als Erklärungsmodell für multivariate Logits im Fall einer polytomen Zielvariablen betrachten. Wir erläutern diese Verallgemeinerung wiederum an dem in Abschnitt (B) eingeführten Beispiel. Die Ausgangstabelle der einzelnen Ausprägungskombinationen für die Merkmale STADT, VAB und BILDUNG findet sich auf S. 133.

Wir betrachten zunächst die logarithmierten Häufigkeiten in den Ausprägungskombinationen $(S_{1j}V_{jk})$, j. abhängige Variable (') und spezifizieren eine Hierarchie von (symmetrischen) log-linearen Modellen, die in logistische (Abhängigkeits-) Modelle transformiert werden können.

Die Suche nach einem den beobachteten Daten hinreichend angepaßten Modell beginnt zunächst mit der Festlegung eines Minimal- (oder Basis-) Modells: Das restriktivste Logit-Modell, in dem alle Haupt- und Interaktionseffekte bzgl. der Variablen BILDUNG gleich Null angenommen werden, entspricht dem log-linearen Modell

BILDUNGSNIVEAU VON SÖHNEN IM KAISERREICH (N = HÄUFIGKEIT IN DER AUSPRÄGUNGSKOMBIATION $S_i V_j B_k$)

STADT	VAB	BILDUNG	N
1	1	1	10
1	1	2	118
1	1	3	68
1	2	1	394
1	2	2	891
1	2	3	251
1	3	1	1580
1	3	2	449
1	3	3	21
2	1	1	24
2	1	2	64
2	1	3	60
2	2	1	1969
2	2	2	633
2	2	3	247
2	3	1	6578
2	3	2	270
2	3	3	19

Ausprägungen der Variablen

STADT: (1) Minden (2) Duisburg;

VAB : (1) Besitz- und Bildungsbürger (2) alter und neuer
Mittelstand
(3) Arbeiter und untere Beamte;

BILDUNG: (1) Volksschule (2) Mittelschule bis Tertia
(3) Sekunda und Prima.

$$\ln \hat{m}_{ijk} = \lambda_0 + \lambda_i^S + \lambda_j^V + \lambda_k^B + \lambda_{ij}^{SV}$$

mit den anzupassenden Randverteilungen (SV) (B). Die Devianz des Minimalmodells, die als Maß für die Gesamtstreuung in den Daten aufgefaßt wird, beträgt 4762. Mit der sukzessiven Einführung von Haupteffekten der Variablen STADT und VAB bzgl. der Variablen BILDUNG nimmt die Devianz ab (s.d. Ergebnisse einzelner Modelltests in der Tabelle auf S. 135). Bezieht man die einzelnen Devianzen auf die Abweichung des Minimalmodells von den beobachteten Daten, können die bereits diskutierten deskriptiven Statistiken zur Beschreibung der Modellanpassung berechnet werden:

$$\text{Erklärte Devianz des fraglichen Modells} = \frac{\text{Dev.d.Basismodells} - \text{Dev.d.fragl. Modells}}{\text{Dev. d. Basismodells}}$$

In unserem Beispiel ergibt erst das Modell mit allen Effekten zweiter Ordnung eine ausreichende Anpassung an die beobachteten Häufigkeiten (erklärte Devianz B = 99,9%). Die Einflußstärke der einzelnen Haupteffekte bzgl. der Variablen BILDUNG ist nun sehr unterschiedlich, wie die relativen Erklärungsanteile (in der Tabelle mit DB bezeichnet) zeigen: Der anteilige Devianzzuwachs, der durch die Vernachlässigung des Effektes X'' (= Einfluß des Vaterberufs auf den Bildungsrang) entsteht, ist mit 63% wesentlich größer als der anteilige Devianzzuwachs von 29%, der durch den Ausschluß des Effektes X''' (= Einfluß der Variablen STADT auf den Bildungsrang) entsteht.

Die mit Hilfe des Programms GLIM ermittelten Koeffizienten zur Beurteilung der Modellanpassung unterscheiden sich zwar nicht von den bereits in Abschnitt B dokumentierten Werten, der Übergang von einem log-linearen Modell zu einem Abhängigkeitsmodell für multivariate Logits im Fall polytomer abhängiger Variablen ist jedoch im Hinblick auf die Umrechnung der geschätzten log-linearen Effekte in Parameter des interessierenden Logit-Modells einfacher. Ein einfaches Beispiel mit einer dichotomen Zielvariablen soll den Übergang vom log-linearen Modell zum Logit-Ansatz verdeutlichen, wobei hier zu besserer Unterscheidung die logistischen Effekte mit γ bezeichnet werden:

Gleichung	A	A(2)	A(3)	A(4)	Y		Summe	Subpopulation
					=1	=0		
π_0	1	0	0	0	n_{11}	n_{10}	n_1	1
$\pi_0 + \pi_2$	2	1	0	0	n_{21}	n_{20}	n_2	2
$\pi_0 + \pi_3$	3	0	1	0	n_{31}	n_{30}	n_3	3
$\pi_0 + \pi_4$	4	0	0	1	n_{41}	n_{40}	n_4	4

MODELLE ZUR ERKLÄRUNG DER BEOBACHTETEN DATEN				
MODELL:	DEVIANZ:	D.F.:	B:	DB:
S+V+B+S.V+S.B+V.B	2.0	4	0.999	-0.00
S+V+B+S.V+S.B	3039.5	8	0.362	-0.63
S+V+B+S.V +V.B	1395.3	6	0.707	-0.29
S+V+B+S.V (Basismodell)	4762.1	10	0.000	-1.00

PARAMETERSCHÄTZUNGEN FÜR DAS HAUPTEFFEKTMODELL		
--- S + V + B + S.V + S.B + V.B ---		
ESTIMATE:	S.E.:	PARAMETER:
1.952	0.20	%GM
1.342	0.12	STADT(2)
4.034	0.19	VAB(2)
5.412	0.19	VAB(3)
2.813	0.19	BILDUNG(2)
2.319	0.20	BILDUNG(3)
0.254	0.12	STADT(2).VAB(2)
0.085	0.12	STADT(2).VAB(3)
-1.939	0.05	STADT(2).BILDUNG(2)
-1.580	0.09	STADT(2).BILDUNG(3)
-2.007	0.19	VAB(2).BILDUNG(2)
-2.796	0.20	VAB(2).BILDUNG(3)
-4.070	0.19	VAB(3).BILDUNG(2)
-6.614	0.25	VAB(3).BILDUNG(3)

Unter Verwendung von "cornered effects" lautet das Regressionsmodell für die vier Subpopulationen:

$$\begin{aligned}
 \text{F.32} \quad \ln(n_{11}/(1 - n_{11})) &= \gamma_0 \\
 \ln(n_{21}/(1 - n_{21})) &= \gamma_0 + \gamma_2 \\
 \ln(n_{31}/(1 - n_{31})) &= \gamma_0 + \gamma_3 \\
 \ln(n_{41}/(1 - n_{41})) &= \gamma_0 + \gamma_4
 \end{aligned}$$

Die Logits $\ln(n_{i1}/(1 - n_{i1}))$ ergeben sich aber aufgrund der logarithmierten Verhältnisse der entsprechenden Häufigkeiten:

$$\begin{aligned}
 \text{F.33} \quad \ln(n_{11}/(1 - n_{11})) &= \ln(\mu_1/\mu_2) = \eta_1 - \eta_2 \\
 \ln(n_{21}/(1 - n_{21})) &= \ln(\mu_3/\mu_4) = \eta_3 - \eta_4 \\
 \ln(n_{31}/(1 - n_{31})) &= \ln(\mu_5/\mu_6) = \eta_5 - \eta_6 \\
 \ln(n_{41}/(1 - n_{41})) &= \ln(\mu_7/\mu_8) = \eta_7 - \eta_8
 \end{aligned}$$

Diese Häufigkeiten sind die abhängige Variable eines log-linearen Modells und daher eine Funktion der log-linearen Parameter in dem Vektor $\underline{\beta}$. Jedes Logit des logistischen Modells ergibt sich somit aus der Differenz der entsprechenden Prädikatoren des log-linearen Modells. Durch Einsetzen in Gleichung F.32 und Vergleich mit Gleichung F.31 erkennt man, in welchem Verhältnis logistische Parameter $\underline{\gamma}$ und log-lineare Parameter $\underline{\beta}$ zueinander stehen:

$$\begin{aligned}
 \text{F. 34} \quad \eta_1 - \eta_2 &= \beta_0 - (\beta_0 + \beta_2^Y) &= -\beta_2^Y &= \gamma_0 \\
 \eta_3 - \eta_4 &= \beta_0 + \beta_2^A - (\beta_0 + \beta_2^A + \beta_2^Y + \beta_{22}^{AY}) &= -(\beta_2^Y + \beta_{22}^{AY}) &= \gamma_0 + \gamma_2 \\
 \eta_5 - \eta_6 &= \beta_0 + \beta_3^A - (\beta_0 + \beta_3^A + \beta_2^Y + \beta_{32}^{AY}) &= -(\beta_2^Y + \beta_{32}^{AY}) &= \gamma_0 + \gamma_3 \\
 \eta_7 - \eta_8 &= \beta_0 + \beta_4^A - (\beta_0 + \beta_4^A + \beta_2^Y + \beta_{42}^{AY}) &= -(\beta_2^Y + \beta_{42}^{AY}) &= \gamma_0 + \gamma_4
 \end{aligned}$$

Die logistischen Effekte entsprechen den um eine Ordnung höheren Effekten des log-linearen Modells. Sie unterscheiden sich lediglich im Vorzeichen, da wir das Verhältnis der ersten Spalte ($Y=1$) relativ zur zweiten Spalte ($Y=0$) betrachten. Das logistische Absolutglied entspricht dem log-linearen Haupteffekt des Merkmals Y . Die logistischen Haupteffekte des Merkmals A entsprechen den jeweiligen log-linearen Interaktionseffekten der Merkmale A und Y ($A*Y$). Wichtig ist, daß man jeweils die um die Zielvariable des logistischen Modells (Y) erweiterten ranghöheren Effekte verwendet. Ferner ist zu beachten, daß wir unter der Verwendung zentrierter Effekte in Abschnitt (B) die Parameter des log-linearen Modells mit dichotomer Zielvariable ver-

doppeln mußten, um die logistischen Effekte des asymmetrischen Ansatzes zu erhalten. Allerdings haben wir auch in diesem speziellen Fall lediglich die Parameter der um die Zielvariable erweiterten log-linearen Effekte verwendet. Während man demnach bei der Reparametrisierung "cornered effects" die Parameter des logistischen Modells direkt aus den log-linearen Effekten ablesen kann, muß man bei der Wahl zentrierter Effekte die jeweiligen log-linearen Parameter zunächst verdoppeln.

Zur Veranschaulichung der Logit-Bildung im GLM-Ansatz gehen wir im folgenden von dem log-linearen Modell

$$\ln \mu_{ijk} = \beta_0 + \beta_i^S + \beta_j^V + \beta_k^B + \beta_{ij}^{SV} + \beta_{ik}^{SB} + \beta_{jk}^{VB} \quad \text{mit } \beta_1^B = \beta_{11}^{SB} = \beta_{j1}^{VB} = 0$$

aus. Die Werte der log-linearen Effekte unter Annahme der Reparametrisierungsbedingung "cornered effects" sind in der zweiten Tabelle auf S. 135 erfaßt. Wir vergleichen jetzt jede Kategorie B,, mit der ersten Kategorie B, :

$$\begin{aligned} \ln (\mu_{ijk} / \mu_{ij1}) &= \eta_{ijk} - \eta_{ij1} \\ &= \beta_0 + \beta_i^S + \beta_j^V + \beta_k^B + \beta_{ij}^{SV} + \beta_{ik}^{SB} + \beta_{jk}^{VB} - \\ &\quad \beta_0 - \beta_i^S - \beta_j^V - \beta_1^B - \beta_{ij}^{SV} - \beta_{i1}^{SB} - \beta_{j1}^{VB} \\ &= \beta_k^B - \beta_1^B + \beta_{ik}^{SB} - \beta_{i1}^{SB} + \beta_{jk}^{VB} - \beta_{j1}^{VB} \end{aligned}$$

Da wir die Reparametrisierungsbedingung "cornered effects" gewählt haben, gilt:

$$\begin{aligned} \eta_{ijk} - \eta_{ij1} &= \beta_k^B + \beta_{ik}^{SB} + \beta_{jk}^{VB}, \quad \text{für } k = 2, 3 \\ &= \gamma_0 + \gamma_i^S + \gamma_j^V, \end{aligned}$$

$$\text{da } \beta_1^B = \beta_{11}^{SB} = \beta_{j1}^{VB} = 0$$

Die multivariaten logistischen Effekte können somit unmittelbar aus den um die Zielvariable BILDUNG erweiterten log-linearen (Interaktions-) Effekten abgelesen werden. Im einzelnen gilt für das Verhältnis:

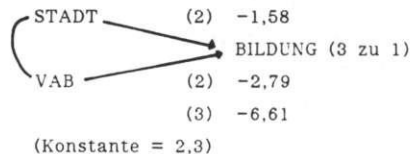
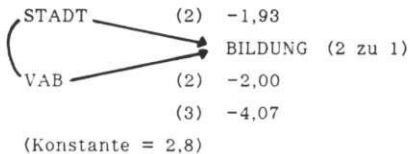
(1) Mittelschule - Tertia zu Volksschule:

$$\ln (\mu_{ij2} / \mu_{ij1}) = \beta_2^B + \beta_{i2}^{SB} + \beta_{j2}^{VB} = \gamma_{20} + \gamma_{2i}^S + \gamma_{2j}^V$$

(2) Sekunda und Prima zu Volksschule:

$$\ln (\mu_{ij3} / \mu_{ij1}) = \beta_3^B + \beta_{i3}^{SB} + \beta_{j3}^{VB} = \gamma_{30} + \gamma_{3i}^S + \gamma_{3j}^V$$

Der Haupteffekt der Variablen STADT bei asymmetrischer Fragestellung entspricht dem Interaktionseffekt STADT.BILDUNG bei symmetrischer Fragestellung; analog entspricht der Haupteffekt der Variablen VAB im logistischen Modell dem Interaktionseffekt VAB.BILDUNG im log-linearen Modell. Das Absolutglied (γ_0) entspricht dem Haupteffekt von BILDUNG bei symmetrischer Fragestellung (β_k^B). Alle anderen Effekte des log-linearen Modells (GM, STADT (2), VAB(2), VAB(3)) werden im logistischen Modell nicht berücksichtigt. Die folgenden Pfaddiagramme zeigen die Einflüsse der Variablen VAB und STADT auf die multivariaten Logits, die sich für die Zielvariable BILDUNG ableiten lassen:



Wie die Modellergebnisse zeigen, ergibt sich nur für die dritte Ausprägung der Variablen VAB eine nennenswerte Abweichung von $6,61 - 4,07 = 2,54$ Punkte. Die einzelnen Variablen wirken auf die Zieldimension in gleicher Richtung. Betrachten wir exemplarisch die Größenordnung der einzelnen Effekte auf das Kategorienverhältnis

(2) Mittelschule - Tertia zu (1) Volksschule

Die Wahrscheinlichkeit, einen Bildungsrang "Mittelschule - Tertia" zu erreichen (im Vergleich zum Bildungsrang "Volksschule"), wird bei Kindern von Arbeitern und unteren Beamten beträchtlich vermindert (um 4,07 Punkte). Bedeutsam sind auch die negativen Effekte STADT(2) und VAB(2). Die Wahrscheinlichkeit für einen Bildungsrang "Mittelschule - Tertia" vs. "Volksschule" wird für Schüler aus Duisburg vermindert. Ähnlich verhält es sich mit dem Effekt der zweiten Ausprägung der Variablen VAB (alter und neuer Mittelstand).

Wie können wir nun die Effekte der einzelnen Variablen auf das (logarithmierte) Verhältnis (3) "Sekunda - Prima" zu (2) "Mittelschule - Tertia" bestimmen? Im folgenden soll gezeigt werden, daß durch leichte Rechnung die von dem Programm GLIM nicht ausgewiesenen logistischen Effekte aus den Parametern des (symmetrischen) log-linearen Modells ermittelt werden können. Hierzu knüpfen wir an die Diskussion unterschiedlicher Reparametrisierungsbedingungen an: Wir hatten bereits festgestellt, daß die Unterschiede zwischen zwei Kategorien unabhängig von der gewählten Reparametrisierungsbedingung eindeutig festliegen. Verdeutlichen wir diese Aussage an dem vorliegenden Beispiel, in dem wir den Zusammenhang zwischen den drei (logarithmierten) Wahrscheinlichkeitsverhältnissen untersuchen:

$$\eta_{ij3} - \eta_{ij2} = \ln \frac{P(B_3)}{P(B_2)} = \ln \frac{P(B_3)/P(B_1)}{P(B_2)/P(B_1)} = \ln \frac{P(B_3)}{P(B_1)} - \ln \frac{P(B_2)}{P(B_1)}$$

Aus dieser Gleichung folgt unmittelbar, daß aus den bereits bekannten logistischen Effekten auf die Wahrscheinlichkeitsverhältnisse $P(B_3)$ zu $P(B_1)$ und $P(B_2)$ zu $P(B_1)$ die gesuchten Effekte auf das Verhältnis der dritten zur zweiten Kategorie der Variablen BILDUNG als spezifische Parameterdifferenzen berechnet werden können:

logistische Effekte auf $\ln P(B_3)/P(B_1)$		logistische Effekte auf $\ln P(B_2)/P(B_1)$		logistische Effekte auf $\ln P(B_3)/P(B_2)$
STADT(2)	-1,58	-	(-1,93)	= +0,35
VAB(2)	-2,79	-	(-2,00)	= -0,79
VAB(3)	-6,61	-	(-4,07)	= -2,54
KONSTANTE	2,3	-	2,8	= -0,5

Die logistischen Effekte auf die jeweils noch verbleibenden multivariaten Logits im Fall polytomer abhängiger Variablen, in denen als Referenzkategorie nicht die erste Ausprägung der Zielvariablen (= Voreinstellung in dem Programm GLIM) gewählt wird, lassen sich also auch aus den um die Zielvariable erweiterten log-linearen Parametern berechnen, in dem die entsprechenden Parameterdifferenzen zwischen jeweils zwei Kategorien der Zielvariablen betrachtet werden.

An dem Effektparameter STADT(2) läßt sich ablesen, daß die Variable STADT auf die Wahrscheinlichkeit eines Bildungsranges "Sekunda - Prima" (relativ zum Bildungsrang "Mittelschule - Tertia") nun in entgegengesetzter Richtung wirkt, wenn gleich dieser Effekt nur von untergeordneter Bedeutung ist. Den stärksten Effekt hat nach wie vor die dritte Ausprägung der Variablen VAB: Die Wahrscheinlichkeit für einen hohen Bildungsrang verringert sich generell für Kinder von Arbeitern und unteren Beamten, auch wenn als Referenzkategorie der mittlere Bildungsrang betrachtet wird.

Im letzten Teil dieses Abschnitts wollen wir uns noch mit der Frage beschäftigen, wie man ein Erklärungsmodell für nominale Zielvariablen mit Interaktionseffekten schätzen und interpretieren kann, denn bislang sind wir lediglich von dem (asymmetrischen) Haupteffektmodell ausgegangen, dem ein bestimmtes (symmetrisches) log-lineares Modell mit den Interaktionstermen STADT.VAB, STADT.BILDUNG und VAB.BILDUNG entspricht. Hauptziel ist es, die Transformation von logistischen Interaktionseffekten in konditionale Haupteffekte im Rahmen des GLM-Ansatzes zu erläutern, da letztere wesentlich anschaulicher zu interpretieren sind.

In dem vorliegenden Beispiel sind wir von dem logistischen Haupteffektmodell mit einer erklärten Devianz von 99% ausgegangen. Dieses Modell ist also bereits eine mehr als hinreichende Beschreibung der beobachteten Daten; Abweichungen vom empirischen

Befund können als zufällig angesehen werden.

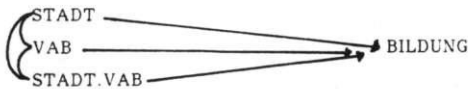
Eine höhere Modellkomplexität erreicht man, wenn der Interaktionsparameter VAB.STADT hinzugefügt wird; dieser Effekt entspricht dem um die Zielvariable BILDUNG erweiterten Interaktionseffekt $\beta_{VAB.STADT.BILDUNG}$ in dem korrespondierenden log-linearen Modell. Inhaltlich hieße dies, daß es für die Prognose der multivariaten Logits durch die unabhängige Variable VAB von Belang ist, ob wir den Schulort MINDEN oder DUISBURG betrachten. Das Modell mit Interaktionseffektparametern bedeutet also, daß der Einfluß der Variablen VAB spezifiziert (oder im Psychologen-Sprachgebrauch: moderiert) wird durch die unabhängige Variable STADT. Da wir nur drei Variablen betrachten, führt die Aufnahme eines Interaktionseffektes dritter Ordnung im log-linearen Ansatz zu einem "saturierten" Modell, da es nun in jedem Fall - per Definition - mit den beobachteten Daten übereinstimmt (d.h. erklärte Devianz = 100%). In unserem exemplarischen Beispiel kommen wir zwar ohne das saturierte Modell und damit dem (logistischen) Interaktionseffekt

STADT.VAB

BILDUNG

aus, wir wollen dennoch im folgenden von einer (hypothetischen) Moderatorwirkung der Variablen STADT ausgehen, um das Vorgehen bei der Modellspezifikation und die Interpretation der Interaktionsparameter an dem bislang verwendeten Beispiel zu veranschaulichen.

Das um den Interaktionseffekt erweiterte Abhängigkeitsmodell



$$\ln (\mu_{ijk} / \mu_{ij1}) = \eta_{ijk} - \eta_{ij1} = \gamma_0 + \gamma_i^S + \gamma_j^V + \gamma_{ij}^{SV}$$

entspricht dem log-linearen Modell

$$\ln \mu_{ijk} = \beta_0 + \beta_j^V + \beta_i^S + \beta_k^B + \beta_{jk}^{VB} + \beta_{ik}^{SB} + \beta_{ji}^{VS} + \beta_{jik}^{VSB}$$

Auf der folgenden Seite sind die Effektparameter des saturierten log-linearen Modells in einer Tabelle zusammengefaßt.

SCHÄTZUNGEN DER PARAMETER IN DEM SATURIERTEN MODELL

--- S + V + B + S.V + S.B + V.B. + S.V.B. ---

ESTIMATE:	S.E.:	PARAMETER:
2.303	0.32	%GM
0.875	0.37	STADT(2)
3.674	0.32	VAB(2)
5.063	0.32	VAB(3)
2.468	0.33	BILDUNG(2)
1.917	0.34	BILDUNG(3)
0.733	0.38	STADT(2).VAB(2)
0.550	0.37	STADT(2).VAB(3)
-1.487	0.40	STADT(2).BILDUNG(2)
-1.001	0.41	STADT(2).BILDUNG(3)
-1.652	0.33	VAB(2).BILDUNG(2)
-2.368	0.35	VAB(2).BILDUNG(3)
-3.726	0.33	VAB(3).BILDUNG(2)
-6.238	0.40	VAB(3).BILDUNG(3)
-0.463	0.41	STADT(2).VAB(2).BILDUNG(2)
-0.624	0.43	STADT(2).VAB(2).BILDUNG(3)
-0.447	0.41	STADT(2).VAB(3).BILDUNG(2)
-0.525	0.52	STADT(2).VAB(3).BILDUNG(3)

mit: STADT(2) = Duisburg;
 VAB(2) = alter und neuer Mittelstand;
 VAB(3) = Arbeiter und untere Beamte;
 BILD(2) = Mittelschule - Tertia;
 BILD(3) = Sekunda - Prima;

Basiskategorien: STADT(1) = Minden;
 VAB(1) = Besitz- und Bildungsbürger;
 BILD(1) = Volksschule.

Da wir von einer asymmetrischen Fragestellung ausgehen, interessieren uns nur die Effekte

BILDUNG(2)	BILDUNG(3)
VAB(2).BILDUNG(2)	VAB(2).BILDUNG(3)
VAB(3).BILDUNG(2)	VAB(3).BILDUNG(3)
STADT(2).BILDUNG(2)	STADT(2).BILDUNG(3)
STADT(2).VAB(2).BILDUNG(2)	STADT(2).VAB(2).BILDUNG(3)
STADT(2).VAB(3).BILDUNG(2)	STADT(2).VAB(3).BILDUNG(3)

für das Verhältnis	für das Verhältnis
"Mittelschule - Tertia"	"Sekunda - Prima"
zu "Volksschule"	zu "Volksschule"

Es zeigt sich, daß die Interaktionen generell die Wahrscheinlichkeit für einen hohen Bildungsrang (relativ zum Volksschulabschluß) vermindern. Die Ausprägungskombination STADT = "Duisburg" und VAB = "alter und neuer Mittelstand" vermindert z. B. die Wahrscheinlichkeit für den Bildungsrang "Mittelschule - Tertia" (im Vergleich zu BILDUNG = "Volksschule") um weitere 0.463 Punkte. Der negative Einfluß der Herkunftsschicht "alter und neuer Mittelstand" (VAB=2) wird also für den Schulort Duisburg noch geringfügig verstärkt. Mit anderen Worten: Der Interaktionsterm quantifiziert den zusätzlichen Einfluß einer bestimmten Ausprägungskombination, der über den Einfluß des jeweiligen Haupteffektes (hier: \nearrow VAB) hinausgeht.

Entsprechend ist auch der Interaktionseffekt

$$\text{STADT(2).VAB(3).BILDUNG(2)} = \text{STADT(2). VAB(3)}$$

(im logistischen Modell mit dem Verhältnis BILD = 2 zu BILD = 1 als abhängige Variable)

zu interpretieren: Verglichen mit dem Haupteffekt der dritten Kategorie der Variablen VAB (-3,726) vermindert sich die Wahrscheinlichkeit für einen Bildungsrang "Mittelschule - Tertia" (relativ zum Volksschulabschluß) um weitere 0,447 Punkte, wenn wir die Stadt Duisburg betrachten. Die Interaktionseffekte sind also generell nur in Verbindung mit den entsprechenden Haupteffekten interpretierbar. Folgende Beispielrechnungen machen diese Überlegungen deutlich:

$$\begin{array}{lcl} \text{Haupteffekt von} & & \\ \text{VAB(2) = -1,65} & \nearrow & \ln \frac{p(\text{BILD}=2)}{p(\text{BILD}=1)} \\ + \text{Interaktionseffekt} & \nearrow & \\ \hline \text{VAB(2).STADT(2) = - 0,46} & & \\ = \text{konditionaler Haupteffekt von} & & \\ \text{VAB(2) für Duisburg (STADT=2) = - 2,11} & & \end{array}$$

$$\begin{array}{lcl}
 \text{Haupteffekt von} & \nearrow & \ln \frac{p(\text{BILD}=3)}{p(\text{BILD}=1)} \\
 \text{VAB(3)} = -3,72 & \searrow & \\
 + \text{ Interaktionseffekt} & & \\
 \hline
 \text{VAB(3).STADT(2)} = -0,44 & & \\
 = \text{ konditionaler Haupteffekt von} & & \\
 \text{VAB(3) für Duisburg (STADT=2)} = -4,17 & &
 \end{array}$$

oder allgemein:

$$\begin{array}{lcl}
 \text{konditionaler Haupteffekt} & = & \text{Haupteffekt} \\
 \text{einer Variablen} & & \text{dieser Variablen} + \text{Interaktionseffekt} \\
 & & \text{dieser Variablen} \\
 & & \text{mit einer weiteren} \\
 & & \text{Kategorialvariablen}
 \end{array}$$

Die Summe von Interaktions- und Haupteffekten ist also gerade gleich dem entsprechenden konditionalen Haupteffekt, der den Effekt einer Variablen in einer bestimmten Teilpopulation quantifiziert (hier: STADT = Duisburg):

$$\begin{array}{lcl}
 \text{VAB(2)} & \nearrow -2,11 & \ln \frac{p(\text{BILD}=2)}{p(\text{BILD}=1)} \\
 & \searrow & \\
 \text{VAB(3)} & \nearrow -4,17 & \\
 & & \text{(für STADT = Duisburg)}
 \end{array}$$

Aus der Gleichung läßt sich schließlich auch ableiten, daß ein konditionaler Haupteffekt und einfacher Haupteffekt dann genau übereinstimmen, wenn der entsprechende Interaktionseffekt gleich Null gesetzt ist.

Da wir die Reparametrisierungsbedingung "cornered effects" mit der jeweils ersten Ausprägung einer Variablen als Referenzkategorie verwenden (''), sind die einfachen Haupteffekte der Variablen VAB identisch mit den korrespondierenden konditionalen Haupteffekten in der Teilgruppe STADT = Minden (Ausprägung '1'):

$$\begin{array}{lcl}
 \text{VAB(2)} & \nearrow -1,65 & \ln \frac{p(\text{BILD}=2)}{p(\text{BILD}=1)} \\
 & \searrow & \\
 \text{VAB(3)} & \nearrow -3,72 & \\
 & & \text{(für STADT = Minden)}
 \end{array}$$

Anstelle des Modells mit Interaktionen hätten wir auch direkt für Minden und Duisburg getrennt ein log-lineares Modell rechnen können, das den konditionalen (stadtspezifischen) Einfluß der Variablen VAB auf die Zielvariable BILDUNG mißt:

PARAMETERSCHÄTZUNGEN FÜR DIE MODELLE ($V \cdot B + V.B$):					
(1) STADT « MINDEN			(2) STADT « DUISBURG		
ESTIMATE [^]	S.E.:	PARAMETER:	ESTIMATE [^]	S.E.:	PARAMETER:
2.303	0.31	%GM	3.178	0.20	%OH
3.671*	0.32	VAB(2)	4.407	0.20	VAB(2)
5.063	0.31	VAB(3)	5.613	0.20	VAB(3)
2.U68	0.33	BILD(2)	0.980	0.2U	BILD(2)
1.917	0.3U	BILD(3)	0.916	0.2j.	BILD(3)
-1.652	0.33	VAB(2).BILD(2)	-2.116	0.24.	VAB(2).BILD(2)
-2.368	0.3U	VAB(2).BILD(3)	-2.992	0.25	VAB(2).BILD(3)
-3.726	0.33	VAB(3).BILD(2)	-U.17U	0.25	VAB(3).BILD(2)
-6.238	0.140	VAB(3).BILD(3)	-6.763	0.33	VAB(3).BILD(3)

Die Berechnung subgruppenspezifischer, log-linearer Modelle hat den Vorteil, daß die konditionalen Haupteffekte direkt bestimmt werden könne, ohne Verrechnungen der einfachen Haupteffekte mit den entsprechenden Interaktionseffekten in einem integrierten Modellansatz durchführen zu müssen. Das Beispiel zeigt aber auch, daß bei dieser Vorgehensweise der direkte Einfluß der Variablen STADT auf die Zielvariable BILDUNG nicht mehr Bestandteil des Modells ist.

Die Interpretationsschwierigkeiten von Interaktionseffekten und das Problem der Ausblendung bestimmter Effekte in konditionalen Abhängigkeitsmodellen kann im Rahmen des GLM-Ansatzes umgangen werden, in dem eine lineare Transformation der ursprünglichen Designmatrix mit den Haupt- und Interaktionseffekten vorgenommen wird (ausführlich s. Arminger 1982, S. 63ff.). Wir erhalten dann eine Design-Matrix mit konditionalen Haupteffekten anstelle des Interaktionseffektes VAB.STADT (im logistischen Modell) und den direkten Effekt der Variablen STADT. Formal, d.h. von der Güte der Anpassung des Modells an die beobachteten Daten her, ist das Modell mit konditionalen Haupteffekten äquivalent zu dem vorangegangenen Modell mit Interaktionseffekten. Das Programm GLIM nimmt diese Transformation dann vor, wenn die Modellspezifikation einen sogenannten "nesting Operator" (/) enthält. Die Anweisung

STADT/(VAB+STADT+VAB.STADT)

bewirkt den Aufbau einer Designmatrix mit den Effekten

S, S.V, S.B, S.V.B

(die Ergebnisse des log-linearen Modells sind in der Tabelle auf der folgenden Seite zusammengefaßt).

Mit anderen Worten: Das log-lineare Modell ($V+B+V.B$) wird innerhalb der Kategorien S , ($i = 1,2$) betrachtet. Inhaltlich entspricht das (symmetrische) log-lineare Modell dem Haupteffektmodell bei asymmetrischer Fragestellung unter Annahme unterschiedlicher subgruppenspezifischer Effekte der Variablen VAB. Bei der Interpretation der Parameter als logistische Effekte ist wieder zu beachten, daß wir nur jeweils die um die Zielvariable BILDUNG erweiterten Effekte verwenden.

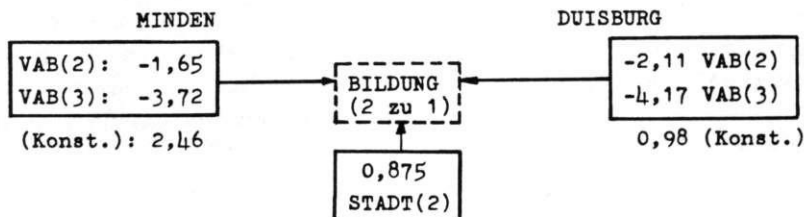
SCHÄTZUNGEN DER PARAMETER IN EINEM SATURIERTEN MODELL MIT KONDITIONALEN EFFEKTEN

--- STADT / (VAB * BILDUNG) ---

--- S + S.V + S.B + S.V.B. ---

ESTIMATE:	S.E.:	PARAMETER:
2.303	0.31	%GM
0.875	0.37	STADT(2)
3.674	0.32	STADT(1).VAB(2)
5.063	0,32	STADT(1).VAB(3)
4.407	0.20	STADT(2).VAB(2)
5.613	0,20	STADT(2).VAB(3)
2.468	0.33	STADT(1).BILDUNG(2)
1.917	0.33	STADT(1).BILDUNG(3)
0.980	0.24	STADT(2).BILDUNG(2)
0.916	0.24	STADT(2).BILDUNG(3)
-1.652	0.33	STADT(1).VAB(2).BILDUNG(2)
-2.368	0.34	STADT(1).VAB(2).BILDUNG(3)
-3.726	0.33	STADT(1).VAB(3).BILDUNG(2)
-6.238	0.40	STADT(1).VAB(3).BILDUNG(3)
-2.116	0.24	STADT(2).VAB(2).BILDUNG(2)
-2.992	0.25	STADT(2).VAB(2).BILDUNG(3)
-4.174	0.24	STADT(2).VAB(3).BILDUNG(2)
-6.763	0.33	STADT(2).VAB(3).BILDUNG(3)

LOGIT-MODELL MIT KONDITIONALEN HAUPTEFFEKTEN



Im Vergleich zu den Parametern des log-linearen Modells mit Interaktionseffekten (s.S. 141) zeigt sich, daß die Konstante (%GM) und der Haupteffekt der Variablen STADT unverändert bleiben: die konditionalen Effekte der Variablen VAB für Minden (STADT = 1) entsprechen den Häufigkeiten von VAB(2) und VAB(3), während die konditionalen Effekte der Variablen VAB für Duisburg (STADT = 2) den Gesamtbeitrag von Haupt- und Interaktionseffekten messen (Beispiele für das Verhältnis BILD=2 zu BILD=1):

	Konditionaler Effekt = Haupteffekt			+	Interaktionseffekt
Minden	VAB(2)	: - 1,65	= - 1,65	+	0
Minden	VAB(3)	: - 3,72	= - 3,72	+	0
Duisburg	VAB(2)	: - 2,11	= - 1,65	+	(-0,46)
Duisburg	VAB(3)	: - 4,17	= - 3,72	+	(-0,45)

Eine nähere Inspektion der einzelnen Parameterschätzungen ergibt die gleichen inhaltlichen Ergebnisse, die wir unter Annahme des logistischen Haupteffektmodells diskutiert haben. Die Schätzwerte für die konditionalen Effekte der Variablen VAB unterscheiden sich zudem nur geringfügig in den einzelnen Subgruppen (Minden versus Duisburg). Das Ergebnis des Stadtvergleichs unter Annahme des saturierten Modells war zu erwarten, denn aufgrund der hinreichenden Anpassung des nicht saturierten Modells (erklärte Devianz = 99%) hatten wir bereits die Modellannahme

"keine differentiellen Effekte der Variablen VAB für Minden und Duisburg"

als mit den beobachteten Daten vereinbar angesehen. Die beobachteten Häufigkeiten in unserer dreidimensionalen Ausgangstabelle enthalten gegenüber den theoretisch erwarteten Häufigkeiten unter dem nichtsaturierten Modell keine weiteren Informationen mehr, die sich in substantiell unterschiedlichen Effekten der Variablen VAB in den Gruppen MINDEN und DUISBURG ausdrücken könnten.

Wie wollen abschließend noch einmal die wichtigsten Ergebnisse zusammenfassen: Mit der Darstellung des log-linearen Modells bzw. des Logit-Modells im Rahmen verallgemeinerter linearer Modelle versuchten wir deutlich zu machen, daß den neueren Verfahren zur multivariaten Analyse qualitativer Merkmale im Vergleich zu den klassischen Verfahren für metrische (abhängige) Merkmale die gleiche Struktur zugrunde liegt, wenn man einmal von den veränderten Randbedingungen bzgl. der Fehlerverteilung und der Verbindungsfunktion absieht. Innerhalb des GLIM-Ansatzes ist auch das Problem der korrekten Behandlung unbesetzter Zellen gelöst. Ein weiterer Vorteil ist der einfache Übergang von log-linearen zu logistischen Modellen durch die Annahmen der Reparametrisierungsbedingung "cornered effects". Darüberhinaus besteht die Möglichkeit, den Einfluß metrischer Merkmale auf eine abhängige nominale Variable in einem Logit-Modell zu untersuchen (s. Arminger 1982, S. 49ff.; Arminger 1982).

G.DIE ANGLEICHUNG VON BEOBACHTETEN (STICHPROBEN-) HÄUFIGKEITEN
EINER MEHRDIMENSIONALEN KONTINGENZTABELLE AN DIE
RANDVERTEILUNGEN DER GRUNDGESAMTHEIT:
DAS VERFAHREN DES "PROPORTIONAL ADJUSTMENT"

Die Analyse einer zweidimensionalen Häufigkeitstabelle mit Hilfe von Abstrom- bzw. Zustromprozenten wirft in unserer Untersuchung der Chancenungleichheit und Mobilität ein gravierendes Problem auf: Die Randverteilungen der betrachteten Merkmale in der Mindener und Duisburger Stichprobe unterscheiden sich von jenen der Grundgesamtheit z.T. erheblich, d.h. es liegt aus erhebungstechnischen Gründen eine im Vergleich zur jeweiligen Grundgesamtheit stark asymmetrische Randverteilung zugunsten der höheren Schulbildungsränge bzw. zugunsten der Oberschicht vor. Das Problem besteht nun darin, daß sowohl die Abstrom- als auch die Zustromprozente von den jeweiligen Randverteilungen der zu analysierenden Tabelle nicht unabhängig sind. Auf der Grundlage der empirischen Häufigkeitstabellen können somit z. B. keine Aussagen über absolute Bildungschancen getroffen werden.

Eine weitere Schwierigkeit tritt bei einer Analyse sozialer Ungleichheit und Mobilität in ihrer zeitlichen Veränderung auf. In der Mobilitätsforschung existiert z.B. eine lange - kontroverse - Diskussion darüber, welche Maßzahlen zur Darstellung der in einer Mobilitätstabelle festgehaltenen Mobilitätsvorgänge auch einen sinnvollen Vergleich der Mobilitätsprozesse in unterschiedlichen Gesellschaften, zwischen mehreren Subpopulationen oder zwischen mehreren Beobachtungszeitpunkten ermöglichen (ausführlich s. Müller 1975; Handl 1983). Die Kritik an den verschiedenen, zur Anwendung empfohlenen Maßzahlen hat vor allem darauf hingewiesen, daß diese die Größe der Abweichungen der beobachteten Unabhängigkeit erwarteten Werten nicht unabhängig von den Strukturen in den Randverteilungen der betrachteten Häufigkeitstabelle erfassen würden. Es sei daher nicht ohne weiteres möglich, Unterschiede in den Kennzahlen zwischen zwei Mobilitätstabellen direkt als Unterschiede der Chancenstruktur zu interpretieren. Dies sei vielmehr nur dann sinnvoll, wenn zuvor die durch die Unterschiedlichkeit der Randverteilungen verursachten Differenzen ausgeblendet würden. Das vorläufige Ergebnis dieser, sich über mehrere Jahre hinwegziehenden Diskussion ist, daß erst mit der Entwicklung log-linearer Modelle zur multivariaten Tabellenanalyse geeignete Methoden zur Verfügung stehen, die diesen Forderungen gerecht werden können (s. vor allem Hauser 1979; Allerbeck/Stork 1980). Das log-lineare Modell sei, so die einhellige Auffassung vieler Autoren, ein Verfahren, das eine Trennung von Struktur- und reinen Ungleichheitseffekten gestattet. "Konzeptuell entsprechen der Strukturmobilität Randverteilungseffekte, der Fluktuationsmobilität die Assoziation zwischen Beruf des Vaters und Beruf des Sohnes" (Allerbeck/Stork 1980, S. 98). In einer methodologisch begründeten Kritik hat Sobel (1983) darauf hingewiesen,

daß zumindest eine Trennung zwischen Struktur- und Ungleichheitseffekten mit Hilfe des log-linearen Modells nicht möglich ist. "...the model of independence does not 'control' for structural mobility (marginal differences) in any sense, and this implies that the association in a mobility table should not be equated (operationally) with the residual concept of circulation mobility. This in turn implies that researchers who are inclined to rescue the 'structure vs. circulation' framework by invoking the third rationale must abandon the log-linear model and revert to the calculation of traditional mobility indices. But this is not a viable option" (Sobel 1983, S. 725). Der u.E. berechtigten Kritik liegt der Gedanke zugrunde, daß aus dem Vergleich zweier hierarchischer log-linearer Modell ($\chi^2_{L(M1)} - \chi^2_{L(M2)}$)

$$M1: \quad \ln \hat{m}_{ijk} = \lambda + \lambda_{ik}^{VT} + \lambda_{jk}^{ST} \quad (\text{mit } \chi^2_{L(M1)}) \text{ und}$$

$$M2: \quad \ln \hat{m}_{ijk} = \lambda + \lambda_{ik}^{VT} + \lambda_{jk}^{ST} + \lambda_{ij}^{VS} \quad (\text{mit } \chi^2_{L(M2)}) ,$$

M1: zugelassen sind lediglich Assoziationen der Randverteilungen der Berufe mit der Zeit.

M2: neben der Veränderung der Berufsstruktur (die Parameter LAMBDA (VT) und LAMBDA (ST)) wird auch eine über die Zeit konstante Assoziation zwischen den Berufsvariablen angenommen,

die resultierende Reduktion der ungeklärten Devianz (relativ zum Basismodell der vollständigen Unabhängigkeit) den Grad der Zirkulationsmobilität nicht quantifiziert. Das log-lineare Modell erlaubt lediglich, die Assoziation zwischen dem Beruf des Vaters und dem Beruf des Sohnes (λ^{VS}) von Randverteilungseffekten bereinigt als über die Zeit invariant anzugeben. Eine weitere Schwäche des log-linearen Modells bei der Anwendung auf zeitbezogene Daten liegt darin, daß Wandel über die Zeit als Abweichung von einem fiktiven Durchschnitt aller Zeitpunkte gefaßt wird (bedingt durch die Reparametrisierung über zentrierte Effekte). Die durch diese Beschränkung hervorgerufenen Probleme bei der Interpretation log-linearer Effekte als Entwicklungstrends sind in einer Arbeit von Wulms (1982) deutlich herausgestellt worden.

Betrachtet man dagegen nur die Struktur etwa der Zustrom- und Abstromprozente in einer Mobilitätstabelle im Zeitvergleich, so ergeben sich weitere Probleme, die wir schon zu Beginn dieses Abschnitts im Zusammenhang mit unseren stark asymmetrischen Stichproben-Randverteilungen dargestellt haben: Die Ergebnisse eines direkten Vergleichs von Abstrom- und Zustromprozenten sind auch von Unterschieden in den Randverteilungen der zu vergleichenden Matrizen abhängig: "Ein Problem beim Vergleich von Mobilitätstabellen mit Hilfe von Abstromprozenten (wie übrigens auch bei Verwendung der Assoziationskoeffizienten oder der Q-Werte) besteht darin, daß diese Maßzahl nicht zwischen Mobilität, die durch die jeweils gegebenen Randvertei-

lungen erzwungen wird und Mobilität, die durch unterschiedliche Interaktionen innerhalb dieser Randverteilungen bedingt ist, trennen kann ... für analytische Zwecke ist es durchaus sinnvoll zu fragen, wie groß denn die Unterschiede in der Mobilitätsstruktur zweier Matrizen sind, wenn man die durch die ungleichen Marginals verursachten Effekte ausblendet und nur jene Bewegungen betrachtet, die auf unterschiedliche Interaktionseffekte zurückgeführt werden müssen" (Handl 1977, S. 503).

Eine Lösung der hier aufgeworfenen Probleme verspricht die Anwendung des erstmals von Mosteller (1968) ausführlich beschriebenen Verfahrens des "proportional adjustment", welches mit dem von Deming/Stephan (1940) entwickelten Iterationsalgorithmus ("iterative proportional fitting") identisch ist. Dieses Verfahren bietet die Möglichkeit, die Randverteilungen einer mehrdimensionalen Häufigkeitstabelle an beliebige Werte anzupassen. Zwar erfolgt in diesem Anpassungsprozeß auch eine Manipulation der absoluten Zellenhäufigkeiten, die innerhalb einer Tabelle gegebene Assoziationsstruktur (abgebildet etwa durch die Werte der Interaktionseffekte eines log-linearen Modells) werden jedoch nicht verändert.

Dieses Verfahren ist u.a. dazu verwendet worden,

- die Randverteilungen von Mobilitätstabellen unterschiedlicher Gruppen oder für verschiedene Zeitpunkte zu standardisieren, um Mobilitätsmuster unbeeinflusst von den durch Unterschiede in den Randverteilungsstrukturen verursachten Effekten zu vergleichen (s. u.a. Handl 1977, S. 493ff.);
- die Informationen historischer Quellen "aus mehreren Einzeltabellen, in denen zum Teil unterschiedliche Variablen in unterschiedlichem Differenzierungsgrad kombiniert sind, in eine einzige 'Supertabelle' zu approximieren, in der alle Variablen aus den Einzeltabellen in ihrer jeweils differenziertesten Aufgliederung kombinatorisch miteinander verknüpft sind" (Arminger/Lijphart/Müller 1981, S. 275).

Wir wollen im folgenden die Grundidee des Verfahrens, das auf die log-lineare Modellbildung zurückgreift, in der Anwendung auf eine fiktive (2x2)-Tabelle illustrieren:

Empirische Häufigkeitstabelle
in der Stichprobe:

	B		
	(1)	(2)	
A	(1) 70	30	100
	(2) 30	70	100
	100	100	200

Randverteilungen in der
Grundgesamtheit:

B	
(1)	(2)
1000	4000
A	
(1)	(2)
2000	3000

Die Gleichung (G.1) beschreibt das saturierte log-lineare Modell für die zweidimensionale Tabelle:

$$G.1 \quad \ln \hat{m}_{ij} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}$$

Die absoluten Besetzungszahlen werden als Funktion dreier Parameter angesehen: Als Funktion der Größenverhältnisse in den beiden Randverteilungen und als Ergebnis der spezifischen Interaktion der beiden Merkmale innerhalb der gegebenen Randverteilungen. Der Grad der Assoziation läßt sich auch mit Hilfe des "odds ratio" ausdrücken:

$$G.2 \quad \alpha = \frac{\hat{m}_{11} \hat{m}_{22}}{\hat{m}_{12} \hat{m}_{21}} = 5,4 \quad \text{(Ein Wert von 1 würde statistische Unabhängigkeit zwischen A und B bedeuten).}$$

Aufgrund der extremen Abweichungen der Stichprobenrandverteilungen von Randverteilungen der Grundgesamtheit liegt es nun nahe, die zweidimensionale Tabelle an die eindimensionalen Randverteilungen anzupassen, wobei der Zusammenhang zwischen A und B erhalten bleiben soll.

Zur Schätzung der neuen Zellenhäufigkeiten wird das Verfahren der iterativen Skalierung von Deming & Stephan angewandt (zur Vorgehensweise im einzelnen s. Arminger 1980, S. 151). Der Iterationsalgorithmus weist nun die notwendige Eigenschaft auf, daß die Häufigkeiten der empirischen Ausgangstabelle durch proportionale Anpassung in mehreren Iterationsschritten nur soweit verändert werden, bis durch adäquate Aufsummierung der Zellenhäufigkeiten alle in der Modellgleichung (G.1) vorgegebenen Randverteilungen reproduzierbar sind, d.h. bis die Werte für λ_i^A und λ_j^B den vorgegebenen Häufigkeiten genügen.

Die Ergebnisse sind auf der folgenden Seite wiedergegeben. Der Zusammenhang zwischen A und B (gemessen in dem "odds ratio") bleibt erhalten. Ein Vergleich der Abstromprozentage der Ausgangstabelle mit den entsprechenden Werten der angepaßten Tabelle zeigt, wie groß die Unterschiede in der Struktur sind, d.h. man gelangt zu entgegengesetzten Schlußfolgerungen, wenn nicht die Strukturen der Randverteilungen der Grundgesamtheit berücksichtigt werden. Stichprobenbedingte Verzerrungen in den eindimensionalen Verteilungen führen bei der Interpretation von Abstromprozenten im Extremfall zu völlig irrealen Ergebnissen. Wir gehen daher in unserer Untersuchung der Chancengleichheit und Mobilität nur von angepaßten mehrdimensionalen Häufigkeitstabellen aus.

Mit Hilfe des Verfahrens des "proportional adjustment" können wir auch im Vergleich zwischen mehreren Zeitpunkten die vorhandenen Struktureffekte durch Angleichung der Randverteilungen konstant halten. Die einzelnen Schritte dieser Technik sollen in zusammenfassenden Stichworten anhand eines Beispiels skizziert werden. ('').

Exemplarisches Beispiel für die Anwendung des Verfahrens
"iterative proportional fitting" nach Deming & Stephan:

Ausgangstabelle:

	(1)	(2)	
(1)	70	30	100
(2)	30	70	100
	100	100	200

Tabelle mit fest vorgegebenen
Randverteilungen:

	(1)	(2)	
(1)	70	30	2000
(2)	30	70	3000
	1000	4000	5000

A x B an neue Randverteilungen
angepaßt:

	(1)	(2)	
(1)	720	1280	2000
(2)	280	2720	3000
	1000	4000	5000

Abstromprozente:

	(1)	(2)	
(1)	70	30	100
(2)	30	70	100
	50	50	100

	(1)	(2)	
(1)	36,0	64,0	100
(2)	9,3	90,7	100
	20	80	100

Zustromprozente:

	(1)	(2)	
(1)	70	30	50
(2)	30	70	50
	100	100	100

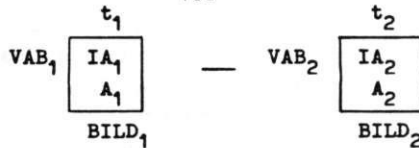
	(1)	(2)	
(1)	72	32	40
(2)	28	68	60
	100	100	100

"Odds ratios":

$$\frac{70 \times 70}{30 \times 30} = 5,4$$

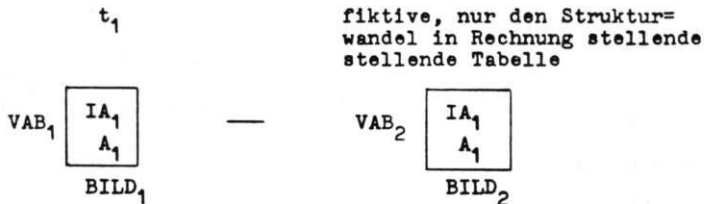
$$\frac{720 \times 2720}{280 \times 1280} = 5,4$$

1. Schritt:

mit IA_1 = Interaktionsstruktur zum Zeitpunkt t_1 ; IA_2 = " " " " t_2 ; A_1 = Abstomprozente " " t_1 ; A_2 = " " " " t_2 ; VAB_1, VAB_2 = Randverteilung der Variablen VAB (t_1, t_2); $BILD_1, BILD_2$ = " " " BILDUNG (t_1, t_2) .

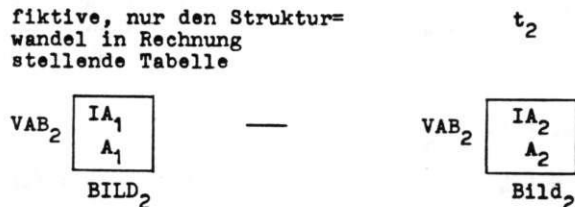
Die Differenzen der Verteilung der Abstomprozente ($A_1 - A_2$) geben Auskunft darüber, in welchem Umfang sich die Chancenstruktur insgesamt (d.h. infolge von Strukturverschiebungen in den Randverteilungen und Veränderungen der Chancenstruktur im engeren Sinne) gewandelt hat.

2. Schritt:



Mit Hilfe des o.g. Iterationsalgorithmus werden die Randverteilungen der Tabelle zum Zeitpunkt t_1 an die zum Zeitpunkt t_2 beobachteten Werte angeglichen, ohne daß die Interaktionsstruktur innerhalb der Tabelle verändert wird. Mit den Differenzen in der Prozentsatzverteilung zwischen geschätzter Tabelle und Ausgangstabelle erhalten wir die Informationen über das Ausmaß des durch die Veränderung der Bildungs- und Berufsstruktur erfolgten Wandels von Chancen.

3. Schritt



Subtrahiert man die Abstromprozentage über diese beiden Tabellen, so folgen daraus die Informationen über die Größe der Veränderung, die allein durch Differenzen in der Chancenstruktur im engeren Sinne verursacht ist, d.h. wir betrachten Bewegungen, die allein auf unterschiedliche Interaktionseffekte zurückgeführt werden können.

Die Berechnung der einzelnen Komponenten soll an dem Beispiel des Wandels der herkunftsspezifischen Bildungschancen zwischen den Zeitintervallen '1830 - 1849' und '1850 - 1885' in Minden illustriert werden.

Zweidimensionale

Ausgangstabellen:

VAB x BILDUNG (t_1, t_2) :

VAB: (1) Besitz- und Bildungsbürger;

(2) alter und neuer Mittelstand;

(3) Arbeiter und untere Beamte;

BILDUNG: (1) Volksschule;

(2) Mittelschule - Tertia;

(3) Sekunda - Prima;

(Die Ausgangstabelle und die geschätzte Tabelle befinden sich auf der folgenden Seite). Zunächst berechnen wir die Differenzen in den beobachteten Abstromprozentsen (M1-M2):

VAB	BILDUNG			T.1: Wandel der herkunfts- spezifischen Bildungschancen im Vergleich der Zeit- intervalle "1830 - 1849" und "1850 - 1885".
	(1)	(2)	(3)	
(1)	-1,32	8,45	- 7,13	
(2)	-6,22	16,99	-10,73	
(3)	-0,6	1,28	- 0,62	

Negative Werte zeigen an, daß sich die Zugangschancen in eine bestimmte Bildungskategorie der Kohorte '1850 - 1885' im Vergleich zur Kohorte '1830 - 1849' absolut gesehen verbessert haben; positive Werte zeigen umgekehrt, daß sich die Zugangschancen in dem zweiten Zeitintervall verschlechtert haben.

Wie sich zeigt, sind für die einzelnen Herkunftsschichten die Zugangschancen zu dem Bildungsrang 'Sekunda - Prima' besser geworden, wenn gleich sie für Arbeiter und untere Beamte nur geringfügig anwachsen. Einen entgegengesetzten Trend beobachten wir für den Bildungsrang 'Mittelschule - Tertia'. Insbesondere für den alten und neuen Mittelstand ist eine deutliche Abnahme der Zugangschancen festzustellen.

Die erste Tabelle zeigt, daß von einer deutlichen Verschiebung der absoluten Bildungschancen ausgegangen werden kann. Berechnen wir daher, in welchem Umfang es allein schon durch die Veränderung der Bildungs- und Berufsstruktur zu einem Wandel der Chancenstruktur gekommen ist (M1 - MG):

BERECHNUNG EINER FIKTIVEN TABELLE, DIE NUR DIE STRUKTURVER-
ÄNDERUNGEN, NICHT JEDOCH DIE VERÄNDERUNG DES INTERAKTIONS-
EFFEKTES IN RECHNUNG STELLT

EINGELESENE MATRIX ZUM ZEITINTERVALL '1830-1849':

VAB	BILDUNG				
	(1)	(2)	(3)		
(1)	0	97	41	138	Absolute Häufigkeiten
(2)	125	625	53	803	
(3)	942	477	6	1425	
	1067	1199	100	2366	

M1:	VAB	(1)	(2)	(3)		
	(1)	0,0	70,29	29,71	100%	Abstrom= prozente
	(2)	15,56	77,83	6,60	100%	
	(3)	66,17	33,47	0,42	100%	

MATRIX ZUM ZEITINTERVALL '1850-1885':

VAB	BILDUNG				
	(1)	(2)	(3)		
(1)	1	47	28	76	Absolute Häufigkeiten
(2)	142	397	113	652	
(3)	641	309	10	960	
	784	753	151	1688	

M2:	VAB	(1)	(2)	(3)		
	(1)	1,32	61,84	36,84	100%	Abstrom= prozente
	(2)	21,78	60,89	17,33	100%	
	(3)	66,77	32,19	1,04	100%	

ANPASSUNG DER RANDVERTEILUNGEN AN DIE IN DER KOHORTE '1850-1885':

MG:	VAB	BILDUNG				
		(1)	(2)	(3)		
	(1)	0,0	48,00	52,00	100%	Geschätzte Matrix unter d. Randvert. t ₂ (Abstrom= prozente)
	(2)	17,18	67,64	15,18	100%	
	(3)	70,03	28,72	1,25	100%	

VAB	BILDUNG		
	(1)	(2)	(3)
(1)	0,0	22,29	-22,29
(2)	-1,62	10,19	- 8,58
(3)	-3,68	4,75	- 0,83

T.2: Differenzen in den Bildungschancen zwischen den Angehörigen der Kohorten '1830 - 1849' und '1850 - 1885' unter der Annahme, daß sich nur die Proportionen der einzelnen Randverteilungen verändert haben.

(Es gilt wiederum: Negative Werte indizieren erhöhte Zugangschancen; positive Werte zeigen demgegenüber eine Verminderung der Zugangschancen an).

Der sich in dieser Tabelle abzeichnende Trend entspricht völlig dem der ersten Tabelle, in der die tatsächlichen beobachteten Chancenänderungen dokumentiert sind. Wenn wir die einzelnen Werte in den Tabellen T.1 und T.2 vergleichen, so finden wir allerdings z.T. erhebliche Größenunterschiede, d.h. allein aufgrund der Strukturveränderungen hätte sich der Zugang zu bestimmten Bildungsabschlüssen in höherem Grad als tatsächlich beobachtbar erhöht oder vermindert. Höhere (negative oder positive) Werte in der Tabelle T.2 bedeuten, daß bei konstanter Chancenstruktur zwischen den Herkunftsschichten allein durch Strukturveränderungen größere Zu- bzw. Abnahmen zu erwarten gewesen wären. Positive oder negative Werte in der Tabelle T.2, die unter den entsprechenden Werten in der Tabelle T.1 liegen, besagen in diesem Zusammenhang, daß der Interaktionseffekt die durch die Strukturveränderungen indizierte Entwicklung verstärkt.

Vergleichen wir etwa den Wert '-29,29' (T.2) mit dem Wert von '-7,13' (T.1), so folgt daraus, daß Kinder von Besitz- und Bildungsbürgern allein aufgrund von Strukturveränderungen in einem höheren Ausmaß als tatsächlich beobachtbar mit dem Abschluß "Sekunda - Prima" ihre Ausbildung beenden hätten müssen. Die Verbesserung der Zugangschancen zu diesem Abschluß hätte also statt der beobachteten Zunahme von 7,13% genau 22,29 Prozentpunkte betragen müssen. Umgekehrt hat sich für den alten und neuen Mittelstand der Zugang zum Abschluß "Mittelschule - Tertia" in einem Umfang verringert, der etwa um 6 Prozentpunkte höher liegt, als aufgrund des Strukturwandels zu erwarten gewesen wäre.

Wir müssen nun noch die Größe des durch die Veränderung des Interaktionseffektes indizierten Wandel ermitteln. Hierzu ist die Differenz zwischen der geschätzten Tabelle MG (Abstromprozente unter der Annahme, es habe allein Strukturveränderungen gegeben) und der Tabelle M2 (tatsächlich beobachtbare Abstromprozente für das zweite Zeitintervall) zu berechnen:

VAB	BILDUNG			Veränderung der relativen herkunftsspezifischen Bildungs- chancen zwischen den betrachteten Zeitintervallen
	(1)	(2)	(3)	
(1)	-1,32	-13,84	15,16	
(2)	-4,6	6,75	- 2,15	
(3)	3,26	- 3,47	0,21	

Mit der Differenz des Interaktionseffektes wird die Veränderung der herkunftsspezifischen Chancenstruktur erfaßt, d.h. wir erhalten das um den Einfluß der Randverteilungsstrukturen bereinigte Muster der Chancenstruktur.

Positive Werte indizieren eine Verringerung der Zugangschancen zu bestimmten Bildungsabschlüssen infolge von Verschiebungen der relativen Chancenstruktur, während negative Werte entsprechend erhöhte Zugänge indizieren. Es wird aus der Tabelle z.B. deutlich, daß sich die Zugangschancen der Besitz- und Bildungsbürger zum Bildungsrang "Sekunda - Prima" relativ verschlechtert haben; Arbeiter und untere Beamte haben im Hinblick auf den Bildungsrang "Mittelschule - Tertia" deutlich gewonnen. Ihre relativen Zugangschancen zum Abschluß "Sekunda - Prima" sind geringfügig zurückgegangen.

Die folgende Gegenüberstellung einzelner Komponenten zeigt noch einmal exemplarisch die Zerlegung der Veränderung der Abstromprozente in der Zeit für die Besitz- und Bildungsbürger:

VAB X BILDUNG	Absolute Veränderung der Chancenstruktur	=	Veränderung der Chancenstruktur = d. Strukturwandel	+	Veränderung der Chancenstruktur = d. Wandel der Interaktion
1 , 1	- 1,32	=	0,0	+	(- 1,32)
1 , 2	8,45	=	22,29	+	(-13,84)
1 , 3	- 7,1	=	-22,29	+	15,16
SUMME:	0,00		0,00		0,00

Zusammenfassend können wir feststellen, daß die Anwendung des Verfahrens der iterativen Skalierung nach Deming & Stephan sowohl die Probleme einer asymmetrischen Stichprobenrandverteilung (relative zur Randverteilung der Grundgesamtheit) löst als auch bei einem Zeitvergleich eine Trennung von Einflüssen der Strukturveränderungen von Verschiebungen der relativen Chancenstruktur ermöglicht. Handl (1983) hat darüberhinaus gezeigt, daß die diskutierte Zerlegung von Abstromprozenten in der Zeit zu inhaltlich adäquateren Ergebnissen führt als etwa alternative Vorgehensweisen, die hier nicht diskutiert wurden.

Abschließend wollen wir noch kurz auf die Quellen der von uns verwendeten EDV-Programme (geschrieben in der Programmiersprache FORTRAN IV) eingehen.

1. Programm IPF zur Anpassung von Kontingenztabelle an vorgebene Randverteilungen:

Quelle: Prof. Dr. G. Armingier
 Fachbereich Wirtschaftswissenschaften
 Universität Gesamthochschule Wuppertal
 Gaußstr. 20
 5600 Wuppertal

2. Programm MATDIFF zur Zerlegung von zeitspezifischen Kontingenztabelle in einzelne Komponenten:

Quelle: Jürgen Sensch
 Fakultät für Soziologie
 Universität Bielefeld
 Universitätsstr.
 4800 Bielefeld 1

H. EDV-PROGRAMME ZUR ANALYSE DISKRETER DATEN

Im letzten Abschnitt möchten wir einen allgemeinen Überblick über die wichtigsten Computerprogramme zur Analyse von diskreten Merkmalen auf der Grundlage von aggregierten Daten bzw. Individualdaten geben. Im folgenden geht es primär darum, durch einen kurzen Softwarevergleich die generellen Eigenschaften der einzelnen Programme herauszustellen sowie ihre Vor- und Nachteile gegeneinander abzuwägen. Wir beschränken die Darstellung auf allgemein zugängliche Software-Pakete; EDV-Programme für sehr spezielle Problemstellungen, die wir jeweils im Anschluß bestimmter Abschnitte genannt haben (z.B. LOGIT und IPF von G. Armingier oder das vom Verfasser erstellte Programm MATDIFF) werden aus der folgenden Diskussion ausgeschlossen.

Die Programme sollten nach Möglichkeit folgende Bedingungen erfüllen:

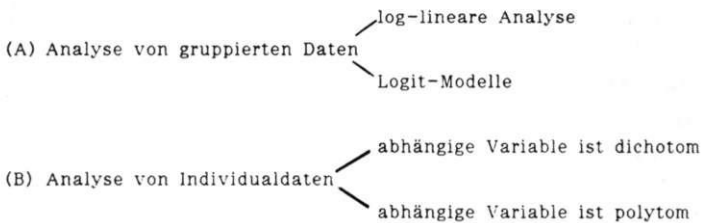
- (1) Sie müssen die jeweils diskutierten Verfahrensschritte umsetzen.
- (2) Die Dateneingabe sollte möglichst variabel sein, wobei nach Möglichkeit auch Datenmodifikationen zugelassen sein sollten.
- (3) Die Prozedurbefehle sollten durchsichtig aufgebaut und ohne spezielle EDV-Kenntnisse vom Benutzer zu bewältigen sein.

- (4) Es sollte eine gute, genügend ausführliche und verständliche Dokumentation vorliegen.

Für einen Software-Vergleich ist es ferner sinnvoll, Kriterien zu entwickeln, welche von den diskutierten Analysemöglichkeiten überhaupt jeweils programmtechnisch umgesetzt sind. Wir gehen daher von folgender Systematik aus:

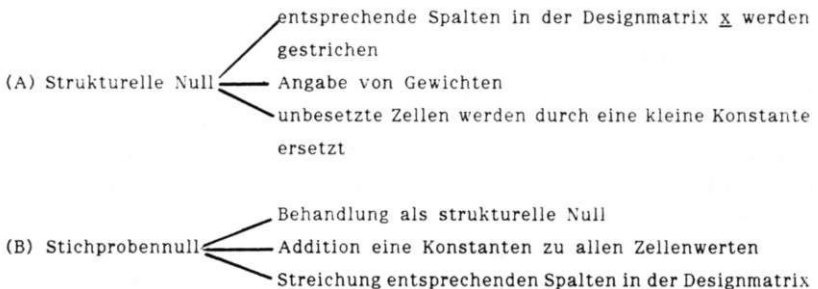
1. Kriterium:

Wir unterscheiden zwischen symmetrischer und asymmetrischer Fragestellung auf unterschiedlichem Aggregationsniveau.



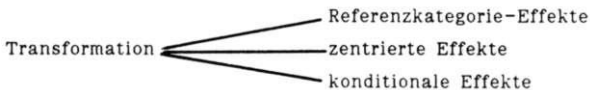
2. Kriterium:

Es gilt zu untersuchen, wie in dem jeweiligen Programm strukturelle oder stichprobenbedingte Nullen behandelt werden können



3. Kriterium:

Als letztes Kriterium verwenden wir den Aufbau der Designmatrix x für den Fall nominaler/ordinaler Merkmale:



Die o.g. Bedingungen und die Umsetzung bestimmter Kriterien konnten am besten von den innerhalb der Programmpakete SPSS, BMDP und SAS vorliegenden Subprogrammen erfüllt werden. Ferner haben wir noch die Programme GLIM (vgl. Baker/Neider 1978) und KALOS (vgl. Röding et al. 1985) berücksichtigt. Die folgende Übersicht umfaßt eine verdichtete Klassifikation der Programme:

PROGRAMM	GRUPPIERTE DATEN		INDIVIDUALDATEN	
	log-lineare A.	Logit-Analyse	abh. Var. dichotom	abh. V. polytom
SPSS ^X (LOGLIN)	+	+	-	-
BMDP (P4F)	+	+	-	-
BMDP (PLR)	-	+	+	-
SAS (FUNCAT)	+	+	+	+
GLIM 3	+	+	+	eingeschränkt
KALOS	-	+	+	+
	"cornered effects"	zentrierte Effekte	konditionale Effekte	
SPSS	+	+	+	
BMDP4F	-	+	-	
SAS	-	+	+	
BMDPLR	+	+	-	
GLIM 3	+	-	+	
KALOS	+	-	+	

PROGRAMM	BEHANDLUNG UNBESETZTER KONTINGENZTABELLENZELLEN		
	Streichung von Spalten	Angabe v. Gewichten	Addition e. Konstanten
SPSS	-	+	+
BMDP4F	-	+	+
BMDPLR	-	-	-
SAS	-	+	-
GLIM 3	+	-	-
KALOS	+	-	-
Abkürzungen: (+) Programm ist speziell für diese Fragestellung zugeschnitten (-) keine Möglichkeit der Berechnung			

Die tabellarische Übersicht erklärt sich weitgehend von selbst, so daß im folgenden nur einige zusätzliche Hinweise gegeben werden sollen.

1. SPSS X: LOGLINEAR

Diese Prozedur untersucht die Modellanpassung, prüft Hypothesen und schätzt Parameter für alle Modelle, die hauptsächlich kategoriale Variablen enthalten. Insofern können mehrdimensionale Kontingenztabellen, Logit-Modelle und logistische Regres-

sionen bearbeitet werden. Das Programm erstellt zunächst auf der Grundlage von Individualdaten anhand der spezifizierten Merkmale eine mehrdimensionale Häufigkeitstabelle. Das Einlesen eines bereits aggregierten Datensatzes ist mit Hilfe der WEIGHT-Anweisung möglich. Somit lassen sich bereits veröffentlichte Tabellen reanalysieren.

Es werden zunächst beobachtete und modellimplizierte Häufigkeiten ausgegeben, Residuen, standardisierte Residuen sowie adjustierte Residuen. Die Ausgabe erfolgt in der Form, daß diese Informationen geschlossen in einer Tabelle integriert sind. Die zahlreichen Angaben innerhalb einer Tabelle gehen hierbei sicherlich zu Lasten einer übersichtlichen Tabellengestaltung. Insbesondere ist zur besseren Lesbarkeit die Vergabe von Etiketten (Variablennamen, Labels für die Merkmalsausprägungen) möglich. Neben der CHI-Quadrat-Statistik zur Beurteilung der Modellanpassungsgüte wird noch die Pearsonsche CHI-Quadrat-Statistik berechnet.

Bei Modellen mit expliziter abhängiger Variable (Logit-Modelle) erstellt die Prozedur zusätzlich Zusammenhangsmaße. Im Rahmen einer asymmetrischen Fragestellung ist es schließlich noch möglich, metrische Kovariaten als Erklärungsgrößen in die Modellgleichung mit aufzunehmen. Da die Analyse lediglich auf dem Niveau einer mehrdimensionalen Kontingenztafel erfolgt, werden als Werte der Kovariaten die subpopulationsspezifischen Mittelwerte berücksichtigt, wobei ein gewisser Informationsverlust in Kauf genommen werden muß.

Mit Hilfe der DESIGN-Anweisung lassen sich die Modellrechnungen durchführen, d.h. diese Anweisung dient der Spezifikation der jeweils anzupassenden Modelle mit den Haupt- und Interaktionseffekten. Eine Schwäche der Prozedur ist es, daß die Tabellendefinition nur implizit über die Modellspezifikation erfolgt, welche wiederum mit Hilfe der Variablennamen vorgenommen wird. Wählt man lange (und somit verständliche) Variablennamen, kann die Modellspezifikation sehr zeitraubend sein.

Für jede Modellspezifikation berechnet das Programm die Regressionskoeffizienten, wobei für nominale/ordinale Variablen sowohl die Effekt- als auch die Dummy-Kodierung als Reparametrisierungsbedingung möglich ist. Liegen stichprobenbedingte oder strukturelle Nullen vor, so können bestimmte Parameter nicht berechnet werden. Durch die Vergabe von Gewichten für die Zellen eines Designs können unbesetzte Zellen jedoch aus der Analyse ausgeschlossen werden, womit keine Probleme mehr bei den Parameterschätzungen auftreten. In derartigen Fällen empfiehlt es sich jedoch, keine Effekt-Kodierung zu wählen, da die Interpretation als Abweichung vom Gesamtmittel verloren geht.

Durch die PRINT-Anweisung werden zusätzliche Ausgaben angefordert, z.B. die Designmatrix, aus der wir sehr übersichtlich die gewählte Reparametrisierungsbedingung ablesen können.

2. BMDP: P4F

Die wesentlichen Leistungen des Programms sind folgende:

- Berechnung und Ausgabe einer Kontingenztafel aus Individualdaten.
- Eingabe einer mehrdimensionalen Tabelle ist möglich.
- Berechnung und Ausgabe von Marginaltabellen beliebiger Dimensionen, d.h. aus einer höherdimensionalen Tabelle lassen sich 'Subtabellen' niedriger Dimensionen erstellen.
- Spezifikation der Tabelle und Modellspezifikation erfolgt separat.
- Die Modellspezifikation ist über frei wählbare einstellige Alpha-Zeichen vorzunehmen.
- Anpassungstests spezifizierter hierarchischer Modelle.
- Berechnung der Modellparameter und der Residuen für spezifizierte Modelle.
- **P4F** analysiert Häufigkeitstabellen, bei denen bestimmte Zellen von der Analyse ausgeschlossen werden können (z.B. strukturelle Nullen); optional kann auf sämtliche Zellen eine Konstante addiert werden (z.B. im Fall von Stichprobenbedingten Nullen).
- Für mehrdimensionale Tabellen können Zeilen, Spalten- und Gesamtprozentage ausgedruckt werden.
- Bildung einer Sequenz von log-linearen Modellen durch schrittweises Hinzufügen oder Entfernen von Interaktionseffekten mit Hilfe einer einzigen Anweisung.

Im Bereich der Präsentation der Ausgabetafel hat BMDP erhebliche Vorteile gegenüber SPSS: Das Programm bietet weitreichende Etikettierungsmöglichkeiten für Variablen und Ausprägungen und liefert reproduktionsreife, sehr übersichtlich gestaltete Tabellen.

Besonders hervorzuheben sind die ausgezeichneten Möglichkeiten der implementierten Suchstrategien zur Modellauswahl, die eine schematische Prüfung aller potentiell möglichen Einzelmodellspezifikationen überflüssig macht (bei vier Merkmalen wären das z.B. über 100 Modellspezifikationen, bei fünf Merkmalen über 2000). In der Regel kann mit maximal drei Läufen ein optimales Modell qualifiziert werden, so daß das Programm extrem effizient einsetzbar ist:

- Mit Hilfe einer einzigen ASSOCIATION-Anweisung werden für jede mögliche Partial- und Marginaltafel sämtliche Tests ausgedruckt. Aus diesen Ergebnissen kann der Benutzer erste Hinweise ableiten, welche Interaktionsterme vernachlässigt werden können bzw. welche Effekte im Zielmodell vertreten sein müssen.
- Für einen zweiten Lauf nutzt man die Möglichkeit der Einzelspezifikation des Modells sowie die automatische Variierung der Modellspezifikation durch Weglassen oder Hinzufügen einzelner Effekt-Komponenten (ADD-/DELETE-Anweisung), wobei die Ergebnisse von CHI2-Differenzen-Tests ausgedruckt werden.

- Im dritten Lauf kann das optimale Modell im einzelnen dokumentiert werden (Berechnung der Effektparameter, Ermittlung der relativen Bedeutsamkeit einzelner Effekt-Komponenten etc.).

Insgesamt betrachtet ist P4F ein Programm zur log-linearen Tabellenanalyse, das in der Mehrzahl der Anwendungen anderen Programmen deutlich überlegen ist. Die Beschreibung von P4F (vgl. Dixon 1985) ist im übrigen aufgrund der klaren und verständlichen Darstellung und der zahlreichen Beispiele von ausgezeichneter Qualität. Als Schwachstellen wären lediglich zu nennen:

- Die Berechnung der Effekte erfolgt nur über die Reparametrisierungsbedingung "zentrierte Effekte".
- Es besteht die Einschränkung der Modellklasse auf hierarchische Modelle. Konditionale Effekte, denen unter inhaltlichen Gesichtspunkten ein großes Interesse zukommt, können nur mit hohem Aufwand berechnet werden.

3. BMDP: PLR

Mit PLR läßt sich die Beziehung zwischen einer dichotomen abhängigen Variablen und einem Satz von unabhängigen Variablen untersuchen. Die abhängige Variable repräsentiert hier binäre Ereignisse (ja oder nein, Erfolg oder Mißerfolg), die mit 0 oder 1 kodiert sein müssen, falls Individualdaten vorliegen. Alternativ kann eine mehrdimensionale Häufigkeitstabelle eingegeben werden. Diese kann entweder aus der Gesamthäufigkeit plus eine der beiden Teilhäufigkeiten Erfolg oder Mißerfolg oder nur aus den beiden Teilhäufigkeiten Erfolg und Mißerfolg bestehen.

Die unabhängigen Variablen können sowohl nominal- als auch intervallskaliert vorliegen.

Die Schätzung der Parameter erfolgt in einer schrittweisen oder einer simultanen logistischen Regression. Bei der schrittweisen Regression wird nach jedem Schritt nach bestimmten (steuerbaren) Kriterien entschieden, welche Variable in das Regressionsmodell aufgenommen bzw. ausgeschlossen werden sollen.

Für jeden Rechenschritt werden folgende statistische Kennwerte berechnet:

- Log-Likelihood-Werte, Änderung gegenüber dem vorherigen Schritt, Quadratsumme der Residuen sowie CHI2-Tests (Güte der Modellanpassung).
- Regressionskoeffizienten, deren Standardfehler sowie der Quotient aus beiden.
- F-Werte, wenn die Auswahl aufgrund der Kovarianzmatrix erfolgt, CHI2-Ergebnisse, wenn die Selektion nach der Maximum-Likelihood-Methode vorgenommen wird.
- Tabelle der logarithmierten Likelihood-Werte.
- Das Programm berechnet optional die aufgrund der Prädiktoren prognostizierten Wahrscheinlichkeiten π und $(1-\pi)$ und stellt sie den beobachteten Werten gegenüber. Je nach dem Aggregationsniveau der Eingabedaten werden entweder die

individuellen Wahrscheinlichkeiten oder aggregierte (Gruppen-) Wahrscheinlichkeiten ausgegeben.

Für den Fall, daß aggregierte Daten in Form einer mehrdimensionalen Kontingenztafel (ohne metrische Prädikatoren) analysiert werden sollen, bietet PLR gegenüber P4F keine gravierenden Vorteile, da log-lineare Modelle in einfacher Weise in asymmetrische Logit-Modelle überführt werden können. PLR zeichnet sich hier lediglich dadurch aus, daß neben zentrierten Effekten auch Referenzkategorieeffekte als Transformationsmöglichkeiten nominaler unabhängiger Merkmale möglich sind.

Vorteile ergeben sich erst, wenn neben nominalen Prädikatoren auch metrische unabhängige Merkmale vorliegen. Insbesondere ist die Möglichkeit der Individualdatenanalyse hervorzuheben, die immer dann sinnvoll ist, wenn

- zahlreiche metrische Prädikatoren vorliegen,
- die Zahl der Untersuchungseinheiten zu gering ist, um eine mehrdimensionale Kontingenztafel mit hinreichend besetzten Zellen zu erzeugen.

Ein zentraler Vorteil ist ferner, daß lediglich eine Individualdatenanalyse richtige Bezugsgrößen (Devianzen) für multiple und partielle Bestimmtheitsmaße liefert, die unabhängig von der Anzahl der erklärenden Merkmale und deren Ausprägungen ist. Bei aggregierten Daten hingegen suggerieren diese Statistiken eine sehr viel bessere Modellanpassung, als in Wahrheit vorliegt, da das saturierte Modell die Daten perfekt reproduziert.

Die praktische Anwendung des Programms wird allerdings durch die Analyse von lediglich dichotomen abhängigen Merkmalen stark eingeschränkt.

4. GLIM

Formalstatistisch betrachtet lassen sich die klassischen Verfahren und die multivariaten Modelle zur Analyse diskreter Daten in ein verallgemeinertes lineares Modell einbetten, wie es von Nelder/Wedderburn (1972) vorgeschlagen und in das GLIM-Programmsystem umgesetzt worden ist. GLIM bietet daher gleichzeitig alle Möglichkeiten klassischer multivariater Analyse (Regressions-, Varianz- und Kovarianzanalyse und darüberhinaus die Möglichkeit mehrdimensionaler Kontingenztabellenanalyse sowie logistische Regressionsanalysen auf der Basis von Individualdaten und aggregierten Daten.

GLIM ähnelt allerdings mehr einer Programmiersprache als einem herkömmlichen Programmpaket, daß für bestimmte Auswertungsprobleme fertige Lösungen in Form von Subprogrammen anbietet. Das GLIM-System ist daher in der Version (Release 3) nicht gerade benutzerfreundlich. Insbesondere setzt die Arbeit mit GLIM neben guten formal-statistischen Kenntnissen auch in der Logik des Programmierens voraus.

Bzgl. der Anwendung von GLIM zur Analyse diskreter Daten sind folgende Punkte bedeutsam:

- Asymmetrische Fragestellungen lassen sich sowohl auf der Basis von aggregierten Daten als auch auf Individualdatenbasis beantworten, wobei in beiden Fällen metrische Prädikatoren zugelassen sind. Im Fall einer abhängigen Variablen mit mehr als zwei Ausprägungen sind jedoch beträchtliche Kunstgriffe für eine Individualdatenanalyse erforderlich.
- Liegen unbesetzte Zellen vor, erfolgt für bestimmte Parameter automatisch "extrinsic aliasing", d.h. nicht berechenbare Parameter werden aus der Designmatrix gestrichen.
- Der automatische Aufbau der Designmatrix erfolgt bei Vorliegen von nominalen Merkmalen analog zur Regressionsanalyse mit Dummy-Variablen, d.h. es werden lediglich Referenzkategorieeffekte berücksichtigt. Allerdings können konditionale Effekte spezifiziert werden.
- Nicht-hierarchische Modelle lassen sich in einfacher Weise berechnen.
- GLIM bietet keinerlei Möglichkeiten der Unterstützung bei der Modellsuche, d.h. die einzelnen Modelle sind jeweils explizit zu spezifizieren.

Um die im GLIM-System gebotenen Möglichkeiten ausschöpfen zu können, ist eine intensive Auseinandersetzung mit den spezifischen Befehlen erforderlich. Eine didaktisch gelungene Darstellung findet man in Andreß (1986).

5. KALOS

Das Programm KALOS wurde von M. Röding, U. Küsters und G. Arminger an der Universität und Gesamthochschule Wuppertal (Fachbereich Wirtschaftswissenschaften) entwickelt und ist direkt vom Zentralarchiv für empirische Sozialforschung, Universität zu Köln, erhältlich (vgl. Röding/Küsters/Armingen 1985). Das Programmsystem dient zur Analyse diskreter abhängiger Variablen mit zwei oder mehr Ausprägungen in Abhängigkeit von qualitativen und quantitativen unabhängigen Variablen. Das Programm berechnet Maximum-Likelihood-Schätzer für multinomiale Logit-Modelle sowohl auf der Basis von Individual- als auch aggregierten Daten. Qualitative Prädiktoren werden durch Dummy-Variablen dargestellt (Referenzkategorieeffekte).

Die korrekte Behandlung von strukturellen und stichprobenbedingten Nullen ist durch automatische Prüfung auf Multikollinearität und anschließender Reparametrisierung gesichert (Streichung bestimmter Spalten in der Designmatrix).

Analog zum GLIM-Programm kann KALOS sowohl interaktiv als auch im Batch verwendet werden.

Die wichtigsten Ausgabestatistiken sind:

- Devianz des Basismodells,
- Anzahl der Modellparameter, die im aktuellen Modell geschätzt werden können und Anzahl der auf Null gesetzten Parameter ("aliasing"),
- Pseudo-Bestimmtheitsmaß R^2 ,

- Parameterschätzwerte und ihre Standardabweichungen,
- optional: Ausgabe der Eingabedaten und der prognostizierten Wahrscheinlichkeiten pro Fall bzw. bei aggregierten Daten pro Gruppe.

Für die Bearbeitung asymmetrischer Fragestellungen ist KALOS hinsichtlich seiner Flexibilität (Kontingenztabellenanalyse/Individualdatenanalyse; dichotome/polytome abhängige Variable) dem BMDP-Programmen deutlich überlegen. Die Druckausgabe ist im Gegensatz zu GLIM klar gegliedert, gut lesbar und umfaßt Ergebnisse, die im Rahmen des GLIM-Systems vom Benutzer erst noch programmiert werden müssen.

Fazit: Mit dem vereinfachten und reduzierten Befehlsvorrat ist es auch einem ungeübten Benutzer möglich, eine Vielzahl von Anwendungen auf einfache Weise umzusetzen. U.E. ist mit KALOS eine interessante Alternative zu PLR oder GLIM gegeben.

6. SAS: FUNCAT

Grundsätzlich vereinigt das SAS-Programm all jene Analysemöglichkeiten für diskrete Variablen, die BMDP und SPSS sowie GLIM und KALOS bieten:

- Neben der Analyse mehrdimensionaler Kontingenztabellen ist eine logistische Regression auf der Basis von Individualdaten mit einer dichotomen wie polytomen abhängigen Variablen möglich.
- Als Eingabedaten erwartet das Programm entweder Rohdaten (d.h. die zu analysierende Tabelle wird vom Programm generiert) oder bereits aggregierte Daten in Form einer Kontingenztafel.
- In der Designmatrix können für nominale Merkmale zentrierte Effekte aufgenommen werden. Ferner können konditionale Effekte spezifiziert werden.
- Stichprobennullen werden wie strukturelle Nullen behandelt (durch Streichung bestimmter Spalten aus der Designmatrix).
- Die Spezifikation der Tabelle erfolgt implizit über die in der Modellspezifikation enthaltenen (vollständigen) Variablennamen (analog zu SPSS).
- Die Aufnahme von unabhängigen metrischen Merkmalen ist auch im Rahmen der Tabellenanalyse möglich (subpopulationsspezifische Mittelwertvektoren analog zu PLR, SPSS oder GLIM bzw. KALOS).
- Bei Logit-Modellen werden die prognostizierten Wahrscheinlichkeiten ausgegeben.

Der Vergleich mit den bereits vorgestellten Programmen zeigt:

- In der Tabellengestaltung schneidet BMDP4F wiederum wesentlich besser ab.
- Ein gravierender Nachteil von FUNCAT ist darin zu sehen, daß innerhalb eines Laufes nur ein Modell bearbeitet werden kann.
- Bei SAS fehlen präzise Bezeichnungen der einzelnen Effekte (es wird lediglich der Gruppennamen ausgegeben), was insbesondere im Fall polytomer Merkmale zu Interpretationsschwierigkeiten führt.

NOTES

- 1.) Selbst die Variable BERUFSSTATUS ist nominalskaliert, da rangmäßig nicht zwischen altem und neuem Mittelstand unterschieden werden kann.
- 2.) Vgl. z.B. Mayer 1978; Featherman/Hauser 1978; lialsey 1980; Goithorpe 1980; Harrigan 1980; Alierbeck/Stork 1980; Müller 1981; Hout 1983.
- 3.) Zu nennen sind u.a. Küchler 1979, Arminger 1979, Langeheine 1980, Hamerle/Tutz 1984, Kemper 1982.
- 4.) Der Wert dieser Statistik stellt ein Maß zur Beurteilung der Güte der Modellanpassung dar (Anpassungsstatistik). Im allgemeinen sind die Anpassungsstatistiken so gebildet, daß die Modellanpassung als um so schlechter zu beurteilen ist, je "größer" der Wert der Anpassungsstatistik ist. Die Größe einer Teststatistik ist jedoch immer nur relativ zu sehen, denn der Wert wird unmittelbar beeinflusst von ihrer Verteilung. Wählt man z.B. die übliche χ^2 - Anpassungsstatistik zur Beurteilung der Modellanpassung, so ist ihr Wert immer nur im Verhältnis zu den Freiheitsgraden der χ^2 - Verteilung zu sehen, d.h. ein χ^2 -Wert von 10 ist relativ klein, wenn das Modell selbst 10 oder mehr Freiheitsgrade hat. Er ist jedoch als relativ groß anzusehen, wenn das Modell nur über 5 oder sogar weniger Freiheitsgrade verfügt. Um diese Abhängigkeit des Wertes der Anpassungsstatistik zu umgehen, wählt man statt des Testwertes selbst den zugehörigen p-Wert. Sei $T = \chi^2$ die ML-Anpassungsstatistik und sei $T(M)$ die konkrete Realisation der Statistik für eine konkrete Modellhypothese M (z.B. das Merkmal A und B ist in der zweidimensionalen Kontingenztafel $A \times B$ voneinander unabhängig), bei einem gegebenen Datensatz, dann ist der p-Wert oder die sog. Überschreitungswahrscheinlichkeit definiert als Wahrscheinlichkeit, eine noch extremere, d.h. noch schlechtere Modellanpassung als die beobachtete Modellanpassung $T(M)$ zu erhalten, wenn das Modell tatsächlich richtig ist, d.h.

$$p = p(M) = P(T \geq T(M)).$$

Der p-Wert eines Modells liegt also im Einheitsintervall und ist selbst eine Zufallsvariable, die auf (0,1) gleichverteilt ist, wenn das Modell bzw. die Hypothese Gültigkeit hat. Im Gegensatz zum absoluten Wert der Anpassungsstatistik selbst gilt hier nun: je kleiner der p-Wert, desto schlechter die Modellanpassung. Bezeichne T des $(1-\alpha)$ -Fraktil der χ^2 -Verteilung (oberer kritischer Wert zum Niveau α , α = Irrtumswahrscheinlichkeit), dann besteht zwischen der Teststatistik T und dem p-Wert der eindeutige Zusammenhang

$$p(M) > \alpha \iff T(M) < T_{1-\alpha}.$$

- 5.) Formal heißt dies, daß \underline{X} nicht von vollem Spaltenrang ist. \underline{X} erfüllt nicht die Bedingung, daß der Spaltenrang von \underline{X} gleich der Anzahl der Spalten von \underline{X} ist, die zur Berechnung der Inversen notwendig ist.
- 6.) Z.B. 4F in dem Programmpaket BMDP, GLIM oder HILOGLINEAR in dem SPSS-Programm
- 7.) Es sei angemerkt, daß wir uns im folgenden auf die Diskussion sogenannter "umfassender Modelle" (vgl. Hamerle/Tutz 1984, S. 422) beschränken wollen. Sie sind dadurch charakterisiert, daß in ihnen sämtliche Haupteffekte enthalten sind. Reduziert man die Parametermenge soweit, daß auch Haupteffekte einer oder mehrerer Variablen vernachlässigt werden, folgen daraus weitere, nicht umfassende Modelle einfacher Struktur. In der empirischen Forschungspraxis sind Modelle dieses Typs allerdings inhaltlich wenig interessant und in den seltensten

Fällen mit dem empirischen Befund verträglich.

- 8.) Dieses Verfahren vermeidet auch eine große Schwäche der diskutierten Signifikanzbetrachtungen. Die statistische Signifikanz hängt sehr wesentlich von der Stichprobengröße ab. Sehr große Stichproben wie hier im Umfang von 13000 lassen auch als substanziell vernachlässigbar erscheinende Modellabweichungen signifikant erscheinen. Reynolds (1977) hat gezeigt, daß bei großen Stichproben nur sehr komplexe Modelle mit den beobachteten Daten verträglich sind, das Signifikanzkriterium als alleinige Entscheidungsgrundlage in diesen Fällen somit versagt. Der Vergleich eines fraglichen Modells mit einem Basismodell ist ein heuristisches Verfahren für große Stichproben. Danach soll die Güte eines Modells daran gemessen werden, ob ein beträchtlicher Anteil der Gesamtvariation durch dieses Modell erklärt werden kann.
- 9.) In diesem Beispiel wären bei Gültigkeit etwa des Modells (S) (BV) die Variablen BILDUNG und VAB miteinander assoziiert, während die Variable STADT weder mit BILDUNG noch mit VAB zusammenhängt. Der Ausdruck "partiell" besagt darüberhinaus, daß der Assoziationsgrad zwischen BILDUNG und VAB für die Ausprägungen STADT = 1 (Minden) und STADT = 2 (Duisburg) als gleich angenommen wird.
- 10.) Diese Transformationen verursachen eine nur unwesentliche Verzerrung der Schätzer (vgl. u.a. Fienberg 1980, Kap. 8). Treten unbesetzte Zellen gehäuft auf, so besteht eine weitere Möglichkeit darin, Variablenkategorien zusammenzufassen oder Variablen ganz aus dem Modell zu eliminieren (Reduktion der Dimensionalität einer Kontingenztafel), um so die Häufigkeiten für die betrachteten Zellen zu vergrößern. Hier ist jedoch zu beachten, daß solche "Verdichtungsprozesse" die Zusammenhangsstruktur zwischen den Variablen beeinflussen können und daher nur mit Bedacht angewandt werden (s. Whittemore 1978).
- 11.) Für die multiple Regression erhalten wir folgende Lösungsgleichungen:

$$P(Y_i = 1) = p_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} \quad ,$$

$$P(Y_i = 0) = 1 - p_i = 1 - (\beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki}) \quad .$$
- 12.) Diese Modelle sind nicht linear in den Variablen, jedoch linear in den Parametern.
- 13.) e ist die Basis des natürlichen Logarithmus.
- 14.) Aus Gründen der Übersichtlichkeit vernachlässigen wir im folgenden den Index (i) zur Kennzeichnung der Untersuchungseinheiten $i = 1, \dots, n$.
- 15.) Die Werte für die geschätzten Wahrscheinlichkeiten lassen sich leicht aus den berechneten Regressionskoeffizienten nach der Prognosegleichung B.1.4 des logistischen Modells ermitteln.
- 16.) Zur algebraischen Äquivalenz eines konditionalen und eines multinomialen Logit-Modells s. Maddala 1982, S.42.
- 17.) Infolge der hohen Fallzahl würde die Kapazität des Programms GLIM bei einer Logit-Analyse mit Individualdaten erheblich überschritten.
- 18.) Zur kritischen Diskussion rekursiver Pfadmodelle für mehrdimensionale Tabellen im Rahmen des GOODMAN-Ansatzes vgl. Langeheine (1982).

- 19.) Einen alternativen Lösungsansatz diskutieren wir in dem folgenden Abschnitt.
- 20.) Zur Vergleichenden Diskussion unterschiedlicher Assoziationsmaße in bivariaten Kontingenztabellen und deren Anwendungsvoraussetzungen s. u.a. Böhning/Kleine/Stegman (1982).
- 21.) Modellspezifikation im GLM-Ansatz:
 - Verteilung = Poisson
 - Verbindungsfunktion = log-linear
- 22.) d.h. jeder erste Haupteffekt und jeder Interaktionseffekt mit der ersten Ausprägung einer Variablen wird gleich Null gesetzt.
- 23.) s. Handl 1983, S. 31ff..

LITERATURVERZEICHNIS

Abkürzungen

AJS	:	American Journal of Sociology
ARS	:	Annual Review of Sociology
AS	:	Angewandte Sozialforschung
ASQ	:	Administrative Science Quarterly
ASR	:	American Sociological Review
BIO	:	Biometrika
BJ	:	British Journal of Mathematical and Statistical Psychology
BM	:	Biometrics
EC	:	Econometrica
EPM	:	Educational and Psychological Measurement
JASA	:	Journal of the American Statistical Association
JER	:	Journal of Economic Review
JMR	:	Journal of Marketing Research
JRSS	:	Journal of the Royal Statistical Society
KZfSS	:	Kölner Zeitschrift für Soziologie und Sozialpsychologie
PB	:	Psychological Bulletin
PM	:	Political Methodology
PS	:	Psychometrika
QQ	:	Quality & Quantity
RER	:	Review of Educational Research
SJS	:	Scandinavian Journal of Statistics
SE	:	Sociology of Education
SM	:	Sociological Methodology. San Francisco
SMR	:	Social Methods & Research
SSR	:	Social Science Research
ZfS	:	Zeitschrift für Soziologie

ALLERBECK, K. R., STORK, H. R. (1980): Soziale Mobilität in Deutschland 1833-1970, in: KZfSS (32), S. 76 - 92.

AMEMIYA, T. (1981): Qualitative Response Models - A Survey, in: JER (21), S. 1483-1536.

ANDREß, H. J. (1986): GLIM. Verallgemeinerte lineare Modelle, Braunschweig.

ARMINGER, G. (1979): Log-lineare Modell zur Analyse des Zusammenhangs zwischen nominalen Variablen, in: HOLM, K. (Hg.): Die Befragung 6, München.

- ARMINGER, G. (1980): Ein Verfahren zur Rekonstruktion mangelhafter Strukturdaten, in: Mitteilungsblatt österreichische Gesellschaft für Statistik und Informatik (10), S. 148 - 164.
- ARMINGER, G., LIJPHART, N., MÜLLER, W. (1981): Die Verwendung log-linearer Modelle zur Disaggregation aggregierter Daten, in: Allgemeines Statistisches Archiv (65), S. 273 - 294.
- ARMINGER, G. (1982): Klassisches Anwendungen verallgemeinerter linearer Modelle in der empirischen Sozialforschung, in ZUMA Arbeitsbericht No. 1982/03, S. 1 - 124, Mannheim.
- ARMINGER, G. (1983a): Multivariate Analyse von qualitativen abhängigen Variablen mit verallgemeinerten linearen Modellen, in: ZfS (12), S. 49 - 64.
- ARMINGER, G. (1983b): Analysis of Qualitative Individual Data and of Latent Class Models with Generalized Linear Models. (Paper for the NATO Advanced Research Workshop on Analysis of Qualitative Spatial Data, Amsterdam).
- ASTON, Chr., WILSON, S. R. (1984): Comment on M.B. Brown and C. Fuchs, "On Maximum Likelihood Estimation in Sparse Contingency Tables", in: Computational Statistics % Data Analysis (2), S. 71 - 77.
- BAKER, R. J., NELDER, J. A. (1978): GLIM Manual. Numerical Algorithm Group. Oxford.
- BIRCH, M. W. (1963): Maximum Likelihood in Three Way Contingency Tables, in: JRSS, Ser. B (25), S. 220 - 233.
- BISHOP, Y.M.M., FIENBERG, S.E., HOLLAND, P.W. (1975): Discrete Multivariate Analysis: Theory and Practice, Cambridge, Mass..
- BLAU, P.M., DUNCAN, O.D. (1967): The American Occupational Structure, New York.
- BÖHNING, P., KLEINE, D., STEGMANN, Chr. (1982): Kreuztabellenanalyse, in Soziale Welt (33), S. 102 - 130.
- DEMING, W.E., STEPHAN, F.F. (1940): On a Least Squares Adjustment of a Sampled Frequency Table when the Expected Marginal Totals are Known, in: Annual of Mathematical Statistics (11), S. 427 - 444.
- DIXON, W. (Hg.) (1985): BMDP Statistical Software, Berkeley.
- DOBSON, A.J. (1983): Introduction to Statistical Modelling. London/New York.
- DUNTON, N.E., FEATHERMAN, D.L. (1980): Gender Differences in Intergenerational and Intragenerational Occupational Mobility. University of Wisconsin, Center of Demography and Ecology.
- FEATHERMAN, D.L., HAUSER, R.M. (1978): Opportunity and Change, New York.
- FIENBERG, St.E. (1980): The Analysis of Cross-Classified Categorical Data, 2nd. ed., Cambridge, Mass..
- FORTHOFER, R.N., LEHNEN, R.G. (1981): Public Program Analysis. A New Categorical Data Approach, Belmont, Calif..
- GOLDTHORPE, J.H. (1980): Social Mobility and Class Structure in Modern Britain, Oxford.
- GOODMAN, L.A. (1965): On the Statistical Analysis of Mobility Tables, in: AJS (70), S. 564 - 585.
- GOODMAN, L.A. (1969a): On the Measurement of Social Mobility: An Index of Status Persistence, in: ASR (34), S. 831 - 850.
- GOODMAN, L.A. (1969b): How to Ransack Social Mobility Tables and Other Kinds of Cross-Classifications, in: AJS (75), S. 1 - 39.
- GOODMAN, L.A. (1973): Causal Analysis of Data from Panel Studies and Other Kinds of Surveys, in: AJS (78), S. 1135 - 1191.
- GOODMAN, L.A. (1978): Analyzing Qualitative / Categorical Data. Cambridge, Mass..
- GOODMAN, L.A. (1979): Multiplicate Models for the Analysis of Occupational Mobility Tables and Other Kinds of Cross-Classification Tables, in: AJS (84), S. 804 - 819.
- GRIZZLE, J.E., ST ARMER, C.F., KOCH, G.G. (1969): Analysis of Categorical Data by Linear Models, in: BM (25), S. 489 - 504.
- HALSEY, A.H., HEATH, A.F., RIDGE, J.M. (1980): Origins and Destinations. Oxford.

- HAMERLE, A., TUTZ, G. (1984): Zusammenhangsanalysen in mehrdimensionalen Kontingenztabellen - das log-lineare Modell, in: FAHRMEIR, L., HAMERLE, A. (Hg.): Multivariate statistische Verfahren. Berlin/New York, S. 473 - 574.
- HANDL, J. (1977): Berufliche Chancen von Frauen - Untersuchungen zur weiblichen Berufsmobilität, in: Soziale Welt (28), S. 493 - 523.
- HARRIGAN, P.J. (1980): Mobility, Elites, and Education in French Society of the Second Empire. Waterloo.
- HAUSER, R.M., FEATHERMAN, D.L. (1977): The Process of Stratification. New York.
- HAUSER, R.M. (1979): Some Exploratory Methods for Modelling Mobility Tables and Other Cross-Classified Data, in: SCHUESSLER, K.F. (Hg.): Sociological Methodology 1980, London.
- HECKMANN, J.J. (1978): Dummy Endogenous Variables in a Simultaneous Equation System, in: EC (46), S. 931 - 959.
- HOLM, K. (1979): Das allgemeine lineare Modell, in HOLM, K. (Hg.): Die Befragung 6. München.
- HOUT, M. (1983): Mobility Tables. Beverly Hills.
- KAWASAKY, S., ZIMMERMANN, K.F. (1981): Measuring Relationships in the Log-Linear Probability Model by Some Compact Measures of Association, in Statistische Hefte (22), S. 82 - 107.
- KEMPER, Fr.-J. (1982): Multivariate Analysen für nominalskalierte Daten. Bonn.
- KÜCHLER, M. (1979): Multivariate Analyseverfahren. Stuttgart.
- LANGEHEINE, R. (1980): Log-lineare Modelle zur multivariaten Analyse qualitativer Daten. München.
- LANGEHEINE, R. (1982): Rekursive Pfadmodelle für multivariate Kontingenztabellen, in: Zeitschrift für Sozialpsychologie (13), S. 2 - 12.
- LEE, L.-F. (1981): Fully Recursive Probability Models and Multivariate Log-Linear Probability Models for the Analysis of Qualitative Data, in: Journal of Econometrics (16), S. 51 - 69.
- LUNDGREEN, P., KRAUL, M., DITT, K. SENSCH, J. (Hg.) (1987): Bildungschancen und soziale Mobilität in der städtischen Gesellschaft des 19. Jahrhunderts. Göttingen.
- MADDALA, G.S. (1982): Limited-Dependent and Qualitative Variables in Econometrics. Cambridge, Mass..
- MAGIDSON J. (1981): Qualitative Variance, Entropy, and Correlation Ratios for Nominal Dependent Variables, in: SSR (10), S. 177 - 194.
- MAYER, K.U., MÜLLER, W. (1971): Progress in Social Mobility Research?, in: QQ (5). S. 141 - 178.
- MAYER, K.U. (1979): Strukturwandel im Beschäftigungssystem und berufliche Mobilität zwischen Generationen, In: Zeitschrift für Bevölkerungswissenschaft (5), S. 267-298.
- MAYER, K.U. (1978): Wandel im Beschäftigungssystem, soziale Mobilität und berufliche Flexibilität, in: Beiträge zur Arbeitsmarkt- und Berufsforschung (30), S. 235-267.
- MCCULLAGH, P., NELDER, J.A. (1983): Generalized Linear Models. London/New York.
- MCFADDEN, D. (1974): Conditional Logit Analysis of Qualitative Behavior, in: ZAREMBKA, P. (Hg.): Frontiers in Econometrics. New York.
- MCROBERTS, H.A., SELBEE, K. (1981): Trends in Occupational Mobility in Canada and the United States: A Comparison, in: ASR (46), S. 406 - 421.
- MOSTELLER, F. (1968): Association and Estimation in Contingency Tables, in: JASA (63), S. 1 - 28.
- MÜLLER, W. (1975): Familie - Schule - Beruf. Opladen
- MÜLLER, W., MAYER, K.U. (1976): Chancengleichheit durch Bildung? Stuttgart.
- MÜLLER, W. (1981): Zum Wandel von Chancenstrukturen in der BRD. VASMA-Arbeitspapier Nr. 18, Mannheim.
- NELDER, J.A., WEDDERBURN, R.W.M. (1972): Generalized Linear Models, in: JRSS, Ser.A (135), S. 370 - 384.
- OPP, K.-D., SCHMIDT, P. (1976): Einführung in die Mehrvariablenanalyse. Reinbek bei Hamburg.

- PULLUM**, Th.W. (1975): Measuring Occupational Inheritance. Amsterdam.
- REYNOLDS**, T. (1977): The Analysis of Cross-Classifications. New York.
- ROCHEL**, H. (1983): Planung und Auswertung von Untersuchungen im Rahmen des allgemeinen linearen Modells. Berlin.
- RÖDING**, M., **KÜSTERS**, U., **ARMINGER**, G. (1985): KALOS. Ein interaktives Programmsystem zur Analyse kategorialer Logitmodelle (Programmbeschreibung). Zentralarchiv für empirische Sozialforschung. Universität zu Köln.
- SCHMIDT**, P., **STRAUSS**, R.P. (1975): The Prediction of Occupation Using Multiple Logit Models, in: International Economic Review (16), S. 471 - 488.
- SCHUBÖ**, W., **UEHLINGER**, H.-M. (1986): SPSS X. Stuttgart/New York.
- SOBEL**, M.E. (1983): Structural Mobility, Circulation Mobility and the Analysis of Occupational Mobility: A Conceptual Mismatch, in: ASR (48), S. 721 - 727.
- SWAFFORD**, M. (1980): Three Parametric Techniques for Contingency Table Analysis: A Nontechnical Commentary, in: ASR (45), S. 664 - 690.
- WHITTEMORE**, A.S. (1978): Collapsibility of Multidimensional Contingency Tables, in: JRSS (B 40), S.328-340.
- WILLMS**, A. (1982): Modernisierung durch Frauenarbeit? In: **PIERENKEMPER**, T., **TILLY**, R. (Hg.): Historische Arbeitsmarktforschung. Göttingen, S. 37 - 71.
- YAMAGUCHI**, K. (1983): The Structure of Intergenerational Occupational Mobility: Generality and Specificity in Resources, Channels, and Barriers, in: AJS (88), S. 718 - 744.